

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher <u>Toby Port</u>	NA Sequence (#) <u>2</u>	STN _____
Searcher Phone # <u>308-3534</u>	AA Sequence (#) <u>1</u>	Dialog _____
Searcher Location _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up <u>7/11</u>	Bibliographic _____	Dr Link _____
Date Completed <u>7/12</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time <u>10</u>	Fulltext _____	Sequence Systems <u>CS</u>
Clerical Prep Time _____	Patent Family _____	WWW/Internet _____
Online Time <u>10</u>	Other _____	Other (specify) _____

From: Pak, Michael
Sent: Monday, July 09, 2001 8:21 PM
To: STIC-Biotech/ChemLib
Subject: 09/276,935 sequence search

Sequence search - 2 month amendment
App. #: 09/276,935
Result format: Paper.
Title: an orphan nuclear receptor
Please search:

Search commercial database and interference.

Search SEQ ID NO:13 and 14
Reverse translate seq id no: 14.

Thanks,

Mike Pak

Michael D. Pak
(703)305-7038
CM-1; Rm. 10E13
AU 1646 - USPTO

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 1E01 TEL: 308-3534

CALL 1E01 TEL: 308-3234
Technical Info. Specialist
1E01 TEL: 308-3234

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2001, 12:18:20 ; Search time 2864.58 Seconds

(without alignments)
11587.664 Million cell updates/sec

Title: US-09-276-935b-13

Perfect score: 2146
Sequence: 1 tgaatatatggtgagagaca.....aggtgatgatctgctg 2146

Scoring table: IDENTITY_MTC

Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*

1: qb_ba1:*
2: qb_ba2:*
3: qb_ba3:*
4: qb_in1:*
5: qb_in2:*
6: qb_in3:*
7: qb_om:*
8: qb_ov:*
9: qb_pat1:*
10: qb_pat2:*
11: qb_ph:*
12: qb_p11:*
13: qb_p12:*
14: qb_p13:*
15: qb_p14:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htqo_hum:*
20: em_htqo_inv:*
21: em_htqo_rod:*
22: em_htq_hum1:*
23: em_htq_hum2:*
24: em_htq_hum3:*
25: em_htq_hum4:*
26: em_htq_hum5:*
27: em_htq_hum6:*
28: em_htq_hum7:*
29: em_htq_hum8:*
30: em_htq_inv1:*
31: em_htq_inv2:*
32: em_htq_other:*
33: em_htq_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	2146	100.0	2146	88	AF061056	AF061056 Homo sapi
2	2128.6	99.2	4448	93	HSJ9936	AF069936 Homo sapi
3	2115.4	98.6	3057	10	E32709	E32709 Novel nucle
4	2086.2	97.2	2905	88	AF084645	AF084645 Homo sapi
5	2006.4	93.5	3243	10	E32708	E32708 Novel nucle
6	1896.6	88.4	4337	93	HSJ9937	AF069937 Homo sapi
7	1866.6	87.0	2802	88	AF084644	AF084644 Homo sapi
8	1374	64.0	1374	10	E32701	E32701 Novel nucle

Db 781 AAAACCTTGACACTACTCTCCATTTCAAGAAATTTCCGGCTGGCAGGGGTGCTTACG 840
 QY 841 aatgagctgagagttgagagagctctgagagagagagagagagagagagagagagag 900
 Db 841 AGTGGCTGGAGCTTGGCAGAGTCTGGCAGGGCCCATCGAGGAAACCTCCCAAGTGG 900
 QY 901 aacacagctccggaagaatctgagctcttgaagagctctctgagctgagagagagat 960
 Db 901 AGCCAGAGTCCGGAAGACATTTGCTCTTGAAGGTTCTCTGTCAGTGGCGGGGAGAGAT 960
 QY 961 ggcagatgctggaatacaaacacccagcagcagcagcagcagcagcagcagcagcagc 1020
 Db 961 GGCAGATGCTGGAATACAAACCCAGCCGAGAGTGGCGGGAAGAGATCTTCTCCGCG 1020
 QY 1021 ctgagccacatgagctgagacatgacacacacacacacacacacacacacacacac 1080
 Db 1021 CTGCCCCCATGGCTGACATGTCACATGTCACATGTCACATGTCACATGTCACATG 1080
 QY 1081 gtcacatcctactgag 1140
 Db 1081 GTCATCTCTCTACTTCCAGGACATTTGCCATGAGGACACAGATCTCTCTGTAAGCGGG 1140
 QY 1141 gcttctgagagctgctcaatgagatgacacacacacacacacacacacacacacacac 1200
 Db 1141 GCTTCTGAGCTGTGTCACTGAGATTCACACAGTGTTCACGCGGAGACTGGAACTGG 1200
 QY 1201 gagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1260
 Db 1201 GAGTGTGGCGCGCT 1260
 QY 1261 ctgagagccatgctgagatctcactacacacacacacacacacacacacacacacac 1320
 Db 1261 CTGAGACCCATGCTGAAATTTCCACTGACATGCTGAGAGAGCTGACCTGATGAGGAG 1320
 QY 1321 catgctgagatgag 1380
 Db 1321 TATGCTCTGATGAGGAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
 QY 1381 cagctgagatgag 1440
 Db 1381 CAGCTGTGTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1441 aatcagagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1500
 Db 1441 AATCGGCCCCAGGCT 1500
 QY 1501 ctccgagagatcaatgctcagcagcagcagcagcagcagcagcagcagcagcagcag 1560
 Db 1501 CTCGGAGAGATCAATGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1560
 QY 1561 ttgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1620
 Db 1561 TTGCTACAGCCCTCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1621 ggtgagac 1680
 Db 1621 GGTGACACCTCCGAG 1680
 QY 1681 agacacatgagac 1740
 Db 1681 AGACAGATGAGACATGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1740
 QY 1741 cctgctatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1800
 Db 1741 CCTGCTATGACAGCTGCTAGAGATTTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAG 1800
 QY 1801 tcaagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1860
 Db 1801 TCACTCTGTAG 1860
 QY 1861 agagacacacag 1920
 Db 1861 AGGACACATGAG 1920

QY 1921 aatcctcagacatcccaataaagatgctcaagatgctgaagagagagagagagagag 1980
 Db 1921 AATCCCTCAGATCCCATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGATAG 1980
 QY 1981 ccatcagagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2040
 Db 1981 CCATCTGGGCTCTATGCCACATACCAGGTTTGTGCTGCTGCTGCTGCTGCTGCTG 2040
 QY 2041 accctcagagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2100
 Db 2041 ACCCTAATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
 QY 2101 tggctcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2146
 Db 2101 TGGCTCCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2146
 RESULT 2
 LOCUS HSAJ9936 4448 bp mRNA PRI 08-SEP-1999
 DEFINITION Homo sapiens mRNA for nuclear hormone receptor PRR1.
 ACCESSION AJ009936
 VERSION AJ009936.1 GI:5852062
 KEYWORDS nuclear hormone receptor; orphan nuclear receptor; PRR1 gene.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE Heard D.J., Holloway, J., Hansen, C., Tommerup, N., Aagaard, L. and Vissing, H. Identification of a novel protein isoform of the human nuclear hormone receptor PKX/SXK and localization to chromosome 3q12.1-13.3 Eur. J. Hum. Genet. In press
 TITLE JOURNAL 2 (bases 1 to 4448)
 AUTHORS Heard, D.J.
 JOURNAL Direct Submission
 Submitted (31-JUL-1998) Heard D.J., Novo Nordisk A/S, Molecular Genetics, Novo Alle, DK-2880, Bagsvaerd, DENMARK
 FEATURES
 source
 1..4448
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="q12-q13.3"
 /tissue_type="liver"
 /tissue_type="intestine"
 /tissue_type="colon"
 1771..3144
 /gene="PRR1"
 1771..3144
 /gene="PRR1"
 /note="isoform with longer A/B domain, translation initiates at CUG-1771 5 fold less efficiently than at CUG-1840"
 /codon_start=1
 /transl_except={pos:1771..1773,aa:Met}
 /product="nuclear hormone receptor PRR1-C"
 /protein_id="CA85490.1"
 /db_xref="GI:5852064"
 /translation="MDPGEYGAKNLPNSRGPANLEVPKESWNHTDFKREJTE
 SVPKPSYNADEVEGGFOICRVCCDKATGVHNVMTGCKGFRKARKKARILCP
 RKGACIEIRKTRROQACRLKRLKLESCKNKMINSDEVERRALIKRKSEKSTGP
 LGVGLTEEDRMMIHELMDAQMKTFTTFSSHKNRILGVLSSGTELESLSAASRE
 AAKMSQVRKDLCSLKVSLLRGEDGSVMYKPPADSGKEIFSLPHMAKSTVFG
 IISFAKVISYRDLPIEDQISLNGAEFLCOLFNNVYNAETGIMPGRI SYTELT
 AGTFQQLLEPLKPHYMLKKIQLHEEYVLMQSLSPRPQVYLQKRVYDQLQDF
 AITTKSYTEKRPQPAHREFLIKIMAMLETLSINAOHTRLNLIGDIIHPATLMKE
 LFGITGS"
 1840..3144
 /gene="PRR1"

CDS

[illegible]

	db	2101	IGFACAGATCTGGT	2117	
RESULT #	4				
AF084645					
LOCUS	AF084645	2905 bp	mRNA	pRI	20-OCT-1998
DEFINITION	Homo sapiens orphan nuclear receptor (PARI) mRNA, complete cds.				
ACCESSION	AF084645				
VERSION	AF084645.1	GI:4769538			
KEYWORDS					
SOURCE					
ORGANISM	human.				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2905)				
AUTHORS	Bertilsson,G., Heldrich,J., Svensson,K., Asmu,M., Jendehog,L., Sydow-Eckman,M., Ohlsson,R., Postlund,H., Blomquist,P. and Berkenstam,A.				
TITLE	Identification of a human nuclear receptor defines a new signaling pathway for cyp1A induction				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12208-12213 (1998)				
REFERENCE	9445350				
FEATURES	2 (bases 1 to 2905)				
FEATURE	Bertilsson,G., Asmu,M., Blomquist,P. and Berkenstam,A. Direct Submission Submitted (19-AUG-1998) cell and Molecular Biology, Medical Nobel Institute, Karolinska Institute, Doktorsringen2, Stockholm 17177, Sweden				
LOCATION/Qualifiers					
1..2905	/organism="Homo sapiens"				
/db_xref="taxon:9606"					
/ISSUE_Type="Liver"					
1..2905	/gene="PARI"				
280..1584	/gene="PARI"				
/product_start=1					
/product_end="orphan nuclear receptor"					
/protein_id="AA04558.1"					
/translation="MEVPRKESNMIADEVRHEPTYESVKSYNADEVGQITRV CGAKGYHHNWEETCKGRFRKRKNMNRKCPDFKCAELIKRIRKVCGRGLRKRI LESCRRKEENSDAVEERRLTKRSSEPELGLDYLTEDPKMIREDMDOMOM THDTTSIRKRNKLPGVLSNGELPENLGNAPREBAKKSVYKDLSTKSLVLRG DGSVNVPKPADSKRELFTLLPIRAKSTYMYRGITSPAKVI SYRPDI EDGISTL KAALVELCYLRPNTEVENAEITWEGVSGTLSDPTAGCDQLLEPMKPHYMLKRL LAEEHYLMALSIESTDPDRGVLDIHRVVVDLOLPFAITLKSTIFCNRPYPARPLEFL IAMTELPLNSAOGTORILRLIGPIHPFPALPDQELGITS"					
BASE COUNT	765 a 727 c 778 g 635 t				
ORIGIN					
Query Match	97.2%	Score 2086.2	DB 88	Length 2905	
Best Local Similarity	99.4%	Pred. No. 0			
Matches 2094	Conservative 0	Mismatches 13	Indels 0	Gaps 0	
UY	40	gaatcatacctctgactagggcggcgaaacgaacgttccttcaatggaat	99		
db	16	gaattcatctgaaatctgcggacgacgaacgacncttccttttaattacgagaar	75		
YY	100	ctcagctcagctcgaagccgaagtgtctaatgaaaagcagaqaatgaactada	159		
db	76	cttgccctcaccctgcacagccaagtcttcacagtgaaaaaacgaanmaataa	135		
OY	160	cctctatcctgaacgaagcagcagctccttgaataagctactcctgaatccttga	219		
db	136	ctcttgtcttgaaacaggccaggcgctctcttgtaaaagtacctcttgatcgat	195		
OY	220	gcctgatgtctcaaatggaacctcagcggcgaaaagatcgagctgaagaaactac	279		
db	196	acttgatctgttccaactgacaccccgaaggcgaagaagtcggaagcaaaaact	255		

[illegible]

DB	Accession	Organism	Length	Accession	Organism	Length
DB	2866	AA036672	2925	DB	2925	AA036672
DB	1501	AF084644	1560	DB	1560	AF084644
DB	2926	AF084644	2985	DB	2985	AF084644
DB	1561	AF084644	1620	DB	1620	AF084644
DB	2986	AF084644	3045	DB	3045	AF084644
DB	1621	AF084644	1679	DB	1679	AF084644
DB	3046	AF084644	3105	DB	3105	AF084644
DB	1681	AF084644	1739	DB	1739	AF084644
DB	3106	AF084644	3165	DB	3165	AF084644
DB	1740	AF084644	1799	DB	1799	AF084644
DB	3166	AF084644	3225	DB	3225	AF084644
DB	1800	AF084644	1839	DB	1839	AF084644
DB	3226	AF084644	3285	DB	3285	AF084644
DB	1860	AF084644	1919	DB	1919	AF084644
DB	3286	AF084644	3345	DB	3345	AF084644
DB	1920	AF084644	1979	DB	1979	AF084644
DB	3346	AF084644	3405	DB	3405	AF084644
DB	1980	AF084644	2039	DB	2039	AF084644
DB	3406	AF084644	3465	DB	3465	AF084644
DB	2040	AF084644	2099	DB	2099	AF084644
DB	3466	AF084644	3525	DB	3525	AF084644
DB	2100	AF084644	2146	DB	2146	AF084644
DB	3526	AF084644	3572	DB	3572	AF084644

UY	2006	ccatccttgcgttcacgaagcttttcattgcaactcatagtccaatacccacc	2065		
Dib	1879	cAACGCTTTTCAGTCTCCTTGACGTCTTTTATCTCGTAATTACTGCCTGCCACT	1938		
OY	2006	tccccctggcttccccctcttccagaacctgatgttgqctccaggacctqlaacctcaacg	2125		
Dib	1939	TCCACCCTGGCTCCCCCTCTCTCCCAAGCTCTTTTGAGGCTCAAGGCTGTAACTAAGCC	1998		
OY	2126	cgagtgcaatgaaatgccatgacgacg	2146		
Dib	1999	CAGGCTCAAGCAATCATCTGGG			
RESULT	B				
E3Z701	E3Z701	1374 bp	INA	PAT	07-FEB-2001
DEFINITION	Novel nuclear receptor protein, its gene and utilization thereof.				
ACCESSION	E3Z701	FJ32701.1	G1:13026808		
VERSION	E3Z701.1	G1:13026808			
KEYWORDS	JP 1999127872-A/2..				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens.				
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	1 (bases 1 to 1374)				
TITLE	Jung, Y.S.S., and Naito,				
JOURNAL	Novel nuclear receptor protein, its gene and utilization thereof				
COMMENT	Patent: JP 1999127872-A 2 18 MAY-1999; JAPAN Tokai-CO INC US Homo sapiens (human) PN JP 1999127872 A/2 PB 18-MAY-1999 PF 07 AUG-1998 JP 1998224172 PK JUN YAMANOOTO, YUTAKA SAITO, TAKAHIRI NAITO, PC C12N15/09,C07K14/72,C07K16/28,C12NI/21,C12P21/02,C12Q1/68, PC G01N33/53. PG G01N33/56//((C12N15)db,(C12R1:91),(C12NI/21,C12P1:19), PC (C12P21/02,C12R1:19), PF C12N15/00,(C12N15/00,C12R1:91) CG CG EF Key Location/Qualifiers FT CDS (1)..(1374). FEATURES Source location/Qualifiers 1..1374 for_gamism "Homo sapiens" /db_xref "taxon":9606"				
BASE COUNT	429 a	375 c	406 g	264 t	
ORIGIN					
Query Match	64.0M; Score 1374; DB 10; Length 1374;				
Hest Local Similarity	100.0M; Prod. No.; 0;				
Matches 1374:	Conservative 0; Mismatch 0; Indels 0; Gaps 0;				
OY	245	gtggaccccgagggagaatcgagagcaagaacttacaccaaagcagcagcagccagcagccagagccaca	294		
Dib	1	GTGGACCTCTGAAGAACAATCTGGAAGCAAAGCAAACTTAGCTTAAGCTCAAGCAAGGCGCA	60		
OY	295	gaagcaaatccgagatcgagacccaaaagaacttaacacaaagcagcagcagccagcatttatccctat	354		
Dib	61	GAAACAAAACCTGGACCTCAACACCTAAACAAACCTGGAACCTATATCTGACTTGTATCAACTG	120		
OY	355	gagcacacagcagatcgattcttggaaagccacatgatcaacgcaatgaggaagcctcagagat	414		
Dib	121	GAGCACACAAAGATCTGTTTTCTGAAAAATCAACTGTAAACGCAATATAGAGAAATCGGAGGT	180		
OY	415	cccacaaatcagccagatgatgatgaggaacagagcagctatccatatccatgatccatgaca	474		
Dib	181	CCCATAATCTCCCGTGTATGTGGAGCAAGCGCAATCTGATTAACTTCAATGTCAATGACAA	240		
OY	475	tatgaagatcgagagagctttttcaagagagccatgaaacagcaagccagctgagtagc	534		

[illegible]

Thu Jul 12 14:23:27 2001

us-09-276-935b-13.rge

Page 21

.

:

.

2

Genome version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2001, 15:10:46 : Search time 150.81 Seconds
(without alignments)
8934.931 Million cell updates/sec

Title: US-09-276-935B-13

Perfect score: 2146
Sequence: 1 tgaatataagttgagagaca.....aagtcacatgagatctgtg 2146

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: N.Geneseq_0601: *
2: /SIDSR/qcadata/geneseq/geneseq/NA1980.DAT: *
3: /SIDSR/qcadata/geneseq/geneseq/NA1981.DAT: *
4: /SIDSR/qcadata/geneseq/geneseq/NA1982.DAT: *
5: /SIDSR/qcadata/geneseq/geneseq/NA1983.DAT: *
6: /SIDSR/qcadata/geneseq/geneseq/NA1984.DAT: *
7: /SIDSR/qcadata/geneseq/geneseq/NA1985.DAT: *
8: /SIDSR/qcadata/geneseq/geneseq/NA1986.DAT: *
9: /SIDSR/qcadata/geneseq/geneseq/NA1987.DAT: *
10: /SIDSR/qcadata/geneseq/geneseq/NA1988.DAT: *
11: /SIDSR/qcadata/geneseq/geneseq/NA1989.DAT: *
12: /SIDSR/qcadata/geneseq/geneseq/NA1990.DAT: *
13: /SIDSR/qcadata/geneseq/geneseq/NA1991.DAT: *
14: /SIDSR/qcadata/geneseq/geneseq/NA1992.DAT: *
15: /SIDSR/qcadata/geneseq/geneseq/NA1993.DAT: *
16: /SIDSR/qcadata/geneseq/geneseq/NA1994.DAT: *
17: /SIDSR/qcadata/geneseq/geneseq/NA1995.DAT: *
18: /SIDSR/qcadata/geneseq/geneseq/NA1996.DAT: *
19: /SIDSR/qcadata/geneseq/geneseq/NA1997.DAT: *
20: /SIDSR/qcadata/geneseq/geneseq/NA1998.DAT: *
21: /SIDSR/qcadata/geneseq/geneseq/NA1999.DAT: *
22: /SIDSR/qcadata/geneseq/geneseq/NA2000.DAT: *
23: /SIDSR/qcadata/geneseq/geneseq/NA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2146	100.0	2146	20	AAZ07997
2	2115.4	98.6	3057	20	AAAX59975
3	2101.4	97.9	3093	20	AAAX78807
4	2075.2	96.7	2910	20	AAAX56242
5	2006.4	93.5	3243	20	AAAX59974
6	1866.6	87.0	2802	20	AAAX56243
7	1866.6	87.0	2850	20	AAAX78808
8	1744.4	81.3	2068	20	AAAX89080
9	1374	64.0	1374	20	AAAX59687
10	1330.2	62.0	1422	20	AAAX59688
11	1305	60.8	1305	20	AAAX59666

ALIGNMENTS

RESULT 1
ID AAZ07997 standard: DNA: 2146 BP.
AC AAZ07997:
DT 17-JAN-2000 (first entry)
DE Human pregnane X receptor (hpxr) encoding DNA.
KW Human; nuclear receptor; pregnane X receptor; PXR; CYP: CYP3A4;
KW cytochrome P-450 mono-oxygenase; drug interaction; hpxr; ss.
XX Homo sapiens.
XX MO9948915-A1.
XX 30-SEP-1999.
XX 26-MAR-1999: 99WD-US06737.
XX 27-MAR-1998: 98US-0079593.
XX (GLAXO) GLAXO GROUP LTD.
XX Kiewer SA, Willson TM:
XX MPI: 1999-601202/51.
XX P-PSDB: AAT42691.
XX New human pregnane X receptor, used to identify specific modulators and
XX agents that induce expression of cytochrome P-450 mono-oxygenase
XX Claim 4: Fig 1A-D; 69pp; English.

12	608.2	28.3	644	20	AAAX59972	SEQ ID 12 of JP11
13	278.8	13.0	1960	19	AAV03126	CDNA encoding rat
14	278.8	13.0	2043	14	AAQ51425	Kat vitamin D re
15	272.8	12.7	1382	20	AAAX34789	Nucleotide sequen
16	272.8	12.7	1463	20	AAAX34788	Nucleotide sequen
17	272.8	12.7	1574	20	AAAX34793	Nucleotide sequen
18	272.8	12.7	3510	20	AAZ07545	Human vitamin D re
19	272.8	12.7	4604	19	AAVA1327	Human vitamin D re
20	271.2	12.6	1399	14	AAQ51424	Human vitamin D re
21	257	12.0	2191	17	AAT36499	Xenopus oiphan rec
22	236	11.0	1071	20	AAV16596	Kat vitamin D re
23	236	11.0	3382	19	AAV18518	CDNA encoding rat
24	236	11.0	3382	20	AAV16606	Kat vitamin D re
25	215	10.0	1534	20	AAAX34790	Nucleotide sequen
26	188.6	8.8	1404	20	AAV16597	Human vitamin D re
27	185.4	8.6	1404	19	AAV03130	CDNA encoding rat
28	181.6	8.5	1280	20	AAAX80215	Mouse nuclear rece
29	172.2	8.0	1450	18	AAT92305	Constitutively act
30	170.6	7.9	1450	14	AAQ46131	Human CAR receptor
31	170.6	7.9	1450	20	AAAX23994	Human CAR receptor
32	168.6	7.9	1361	20	AAAX24003	Mouse CAR receptor
33	108.4	5.1	468	20	AAAX80217	Mouse nuclear rece
34	103.6	4.8	1622	21	AAQ98948	Human pancreatic c
35	97.6	4.5	1898	18	AAT79634	DNA encoding human
36	97.6	4.5	1979	17	AAT27616	Human foetal lung
37	97.6	4.5	2030	15	AAQ63134	Human recombinant
38	97.6	4.5	2030	17	AAT18996	Human steroid rece
39	97.6	4.5	2030	17	AAT30031	NR receptor poten
40	97.2	4.5	1813	16	AAU88760	Human ubiquitously e
41	91.4	4.3	2928	13	AAQ29338	KAR-alpha gene, S
42	91.4	4.3	2928	20	AAV64991	Human KAR-alpha cD
43	91.4	4.3	2940	10	AAV90124	DNA of clone pKAR
44	91.4	4.3	3036	13	AAQ28334	myl/KAR-alpha fusi
45	91.4	4.3	3036	19	AAV20474	Human PML/KARalpha

XX The insert ion provides an isolated human nuclear receptor (designated
 CC premature X receptor, PXR) that binds to a cytochrome P-450 mono-oxygenase
 CC (CYP) promoter. The hpxr is used to identify its specific modulators,
 CC and compounds that induce CYP1A4 expression (i.e., to identify drug
 CC interactions, since CYP1A4 is involved in many biotransformations of
 CC drugs). The modulators are potentially useful for associating particular
 CC diseases and conditions with PXR and for treating such conditions.
 CC Antibodies raised against hpxr can be used for determination and
 CC purification of hpxr. The present sequence represents a DNA encoding the
 CC hpxr.

XX Sequence 2146: bp: 509 A; 594 C; 605 G; 439 T; 0 other:

Query Match: 100.0%; Score 2146; 108 20; Length 2146;
 Best Local Similarity: 100.0%; Ident. No. 0;
 Matches 2146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

UY 1 TGAATATAGTGAAGACAAATTTGATCATCTCGAGGAATCATACCATGACTAG 60
DB 1 TGAATATAGTGAAGACAAATTTGATCATCTCGAGGAATCATACCATGACTAG 60
UY 61 GACGGAAGAGGAGGAGCTGCTTACTTCGATGGAGATCTCGGCTCAAGCTG 120
DB 61 GACGGAAGAGGAGGAGCTGCTTACTTCGATGGAGATCTCGGCTCAAGCTG 120
UY 121 AAGATTCAGCATGAGGAAGAGAGGATTAAGCTAATACCTGTCCTGAACAGGAG 180
DB 121 AAGATTCAGCATGAGGAAGAGAGGATTAAGCTAATACCTGTCCTGAACAGGAG 180
UY 181 CAGTCTCTTGAATAGAGTACTTATCGATCTTTACACGAGATGATTCAGATGAG 240
DB 181 CAGTCTCTTGAATAGAGTACTTATCGATCTTTACACGAGATGATTCAGATGAG 240
UY 241 CCGAGGAGGAGATCGAGGAGAGGATACCAACAGGAGTCAGAGGAGGAGAGGA 300
DB 241 CCGAGGAGGAGATCGAGGAGAGGATACCAACAGGAGTCAGAGGAGGAGAGGA 300
UY 301 AAGCTGAGATGAGGAGGAGAGGAGTGGAGGAGTGTGAGTCTGTAGAGTGAAG 360
DB 301 AAGCTGAGATGAGGAGGAGAGGAGTGGAGGAGTGTGAGTCTGTAGAGTGAAG 360
UY 361 AAGAGATCTGTTCTCGAGAGGAGGAGTGTCAAGGAGATGAGGAGTCGAGAG 420
DB 361 AAGAGATCTGTTCTCGAGAGGAGGAGTGTCAAGGAGATGAGGAGTCGAGAG 420
UY 421 AATCTGAGATGATGAGGAGGAGGAGGAGTGTGATTCAGATGATGAGTGTGAG 480
DB 421 AATCTGAGATGATGAGGAGGAGGAGGAGTGTGATTCAGATGATGAGTGTGAG 480
UY 481 GATATGAGGAGGAGTCTTCTGAGGAGGAGTGAATGAGAGGAGGAGTGAAGTCT 540
DB 481 GATATGAGGAGGAGTCTTCTGAGGAGGAGTGAATGAGAGGAGGAGTGAAGTCT 540
UY 541 CAGGAGAGGAGGAGTGTGAGATCAAGGAGAGGAGGAGGAGGAGGAGGAGGAG 600
DB 541 CAGGAGAGGAGGAGTGTGAGATCAAGGAGAGGAGGAGGAGGAGGAGGAGGAG 600
UY 601 CCGAGATGAGTGAAGAGGAGGAGTGAAGAGGAGATGATCATCTCGAGGAGGAG 660
DB 601 CCGAGATGAGTGAAGAGGAGGAGTGAAGAGGAGATGATCATCTCGAGGAGGAG 660
UY 661 GAGAGGAGGAGTGTGATCAAGGAGAGGAGATGAGGAGGAGGAGTCAAGGAGT 720
DB 661 GAGAGGAGGAGTGTGATCAAGGAGAGGAGATGAGGAGGAGGAGTCAAGGAGT 720
UY 721 GTCAGAGGAGTGAAGAGGAGGAGGAGTGTGATCAAGGAGTGAAGGAGTGAAG 780
DB 721 GTCAGAGGAGTGAAGAGGAGGAGGAGTGTGATCAAGGAGTGAAGGAGTGAAG 780
UY 781 AAGAGATCTGAGGAGGAGTGTGATTCAGATGATGAGGAGTGTGAGTGAAG 840
DB 781 AAGAGATCTGAGGAGGAGTGTGATTCAGATGATGAGGAGTGTGAGTGAAG 840

```

```

DB 781 AAGAGATCTGAGGAGGAGTGTGATTCAGATGATGAGGAGTGTGAGTGAAG 840
UY 841 AATGAGGAGGAGTGAAGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
DB 841 AATGAGGAGGAGTGAAGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
UY 901 AAGGAGGAGGAGGAGTGTGAGTCTTGAAGTGTGATGAGGAGGAGGAGGAGGAG 960
DB 901 AAGGAGGAGGAGGAGTGTGAGTCTTGAAGTGTGATGAGGAGGAGGAGGAGGAG 960
UY 961 GAGGAGATCTGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
DB 961 GAGGAGATCTGAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
UY 1021 CCGGAGGAGGAGTGAAGATCAAGTGAAGTGTGAGGAGGAGTGAAGTGTGAGGAG 1080
DB 1021 CCGGAGGAGGAGTGAAGATCAAGTGAAGTGTGAGGAGGAGTGAAGTGTGAGGAG 1080
UY 1081 GTCATCTCTACTTCAGGAGTGTGAGGAGGAGGAGTGTGAGGAGGAGGAGGAG 1140
DB 1081 GTCATCTCTACTTCAGGAGTGTGAGGAGGAGGAGTGTGAGGAGGAGGAGGAG 1140
UY 1141 GCTTTGAGGAGTGTGAGGAGTGAAGATTCAGGAGGAGTGTGAGGAGGAGGAG 1200
DB 1141 GCTTTGAGGAGTGTGAGGAGTGAAGATTCAGGAGGAGTGTGAGGAGGAGGAG 1200
UY 1201 GAGGATGAGGAGGAGTGTGATCTGATGAGGAGGAGTGTGAGGAGGAGTGTGAG 1260
DB 1201 GAGGATGAGGAGGAGTGTGATCTGATGAGGAGGAGTGTGAGGAGGAGTGTGAG 1260
UY 1261 CAGGAGGAGGAGTGTGAGGAGTGAAGTGAAGTGTGAGGAGGAGGAGGAGGAG 1320
DB 1261 CAGGAGGAGGAGTGTGAGGAGTGAAGTGAAGTGTGAGGAGGAGGAGGAGGAG 1320
UY 1321 TATGAGTGTGAGGAGGAGTGTGATCTGAGGAGGAGGAGTGTGAGGAGGAG 1380
DB 1321 TATGAGTGTGAGGAGGAGTGTGATCTGAGGAGGAGGAGTGTGAGGAGGAG 1380
UY 1381 CAGGATGATGAGGAGGAGTGAAGGAGGAGTGAAGTGTGAGGAGGAGGAGGAG 1440
DB 1381 CAGGATGATGAGGAGGAGTGAAGGAGGAGTGAAGTGTGAGGAGGAGGAGGAG 1440
UY 1441 GATGAGGAGGAGGAGTGTGAGGAGTGTGATCTGAGGAGGAGGAGTGTGAGGAG 1500
DB 1441 GATGAGGAGGAGGAGTGTGAGGAGTGTGATCTGAGGAGGAGGAGTGTGAGGAG 1500
UY 1501 CCGGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAG 1560
DB 1501 CCGGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAG 1560
UY 1561 TTTGATGAGGAGGAGTGTGAGGAGGAGTGTGAGGAGGAGGAGGAGGAGGAG 1620
DB 1561 TTTGATGAGGAGGAGTGTGAGGAGGAGTGTGAGGAGGAGGAGGAGGAGGAG 1620
UY 1621 GATGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAG 1680
DB 1621 GATGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAG 1680
UY 1681 AAGGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAG 1740
DB 1681 AAGGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAG 1740
UY 1741 CCGGATGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAG 1800
DB 1741 CCGGATGAGGAGGAGTGTGAGGAGGAGGAGGAGTGTGAGGAGGAGGAGGAG 1800
UY 1801 TCAATCTGATGAGGAGGAGTGTGAGGAGGAGGAGTGTGAGGAGGAGGAGGAG 1860
DB 1801 TCAATCTGATGAGGAGGAGTGTGAGGAGGAGGAGTGTGAGGAGGAGGAGGAG 1860
UY 1861 AAGGAGGAGGAGTGTGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAGGAG 1920
DB 1861 AAGGAGGAGGAGTGTGAGGAGGAGGAGTGTGAGGAGGAGGAGGAGGAGGAG 1920

```



```

OY 1921 aacccttaagctcccaacaaatgctcaaggtgtgtgaaaggaaccaaggaaccaagatag 1980
Db 1921 aatcccttagatcccaactaaagtgtcaaggtgtgtgaaaggaaccaaggaaccaagatag 1980
OY 1981 ccactctgggtctatggtccacatacccaacagttgtgtcctctctctgaagctttcaatgct 2040
Db 1981 ccactctgggtctatggtccacatacccaacagttgtgtcctctctctgaagctttcaatgct 2040
OY 2041 acccttaatagctctgtctctcccaattcccaatctgtcttccctctctcccaagctcttg 2100
Db 2041 accttaataagctctgtctctcccaattcccaatctgtcttccctctctcccaagctcttg 2100
OY 2101 tgggtccagagctctgaactatcatcggaaggtgagatgagatctgtgag 2146
Db 2101 tgggtccagagctctgaactatcatcggaaggtgagatgagatctgtgag 2146

RESULT 2
AAK59975
ID AAK59975 standard; DNA: 3057 BP.
AC AAK59975:
XX
XX
XX 04-AUG-1999 (first entry)
XX
XX
XX DNA encoding an intranuclear receptor protein.
DE
XX
XX Human: intranuclear receptor protein; drug development; diagnosis;
XX treatment; ss.
XX
XX Homo sapiens.
XX
XX JPI1127872-A.
XX
XX 18-MAY-1999.
PD
XX
XX 07-AUG-1998: 98JP-0224172.
XX
XX 11-AUG-1997: 97JP-0230335.
PK
XX
XX (MITSUBI) JAPAN TOBACCO INC.
PA
XX
XX WPI: 1999-350330/30.
DR
XX
XX P-PSDB: AAY16035.
DR
XX
XX
XX
XX New intranuclear receptor protein - useful for drug development and
XX diagnosis and treatment of disease
XX
XX
XX
XX
XX Claim 10: Page 35-37: 38pp; Japanese.
XX
XX
XX The present sequence encodes a human intranuclear receptor protein.
XX The nucleic acid sequence was isolated from a human adult cDNA
XX library using a swissfish ANO2 derived probe. The protein can
XX be used for the development of drugs and diagnosis and treatment
XX of various diseases.
XX
XX
XX Sequence 3057 BP: 792 A; 751 G; 813 C; 701 T; 0 other;

Query Match 98.6%; Score 2115.4; DB 20; Length 3057;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2116; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 30 catatccggaggaatcatctaactatgactagaagcggaggaagagaaacatcttactt 89
Db 1 catatccggaggaatcatctaactatgactagaagcggaggaagagaaacatcttactt 60
OY 90 caatggaaatccgggccttagccttgcgaagccaaqgtgttcaagtgaaagaaagaaagaa 149
Db 61 caatggaaatccgggccttagccttgcgaagccaaqgtgttcaagtgaaagaaagaaagaa 120
OY 150 taagctaaatctctgtccttgaacaaagcagcagcgtcttctgttaaagctactccttga 209

```

Db	121	taagttatattctctgctcttgaaaaagccagcgctctcttctgtuaagttatcttttcttgc	180
QY	210	gattcctttgcaccgaatttcttcaaaqttgagaccaccagggagaaatctctgaacaaatact	259
Db	181	gattcctttgcaccgaatttcttcaaaqttgagaccaccagggagaaatctctgaacaaatact	240
QY	270	accaccagacggtccaaaggcccaagaagcaaaccttgaggttaagaccaccaagaatactg	329
Db	241	accaccagacggtccaaaggcccaagaagcaaaccttgaggttaagaccaccaagaatactg	300
QY	330	gaaccattctacttttgaacacttgagacacacaaatcttcttgaaagcccaatgt	389
Db	301	gaaccattctacttttgaacacttgagacacacaaatcttcttgaaagcccaatgt	360
QY	390	caacgcgaatgaggaagtctggaagttccccaattctgcctgtatatttggaacaaagccac	449
Db	361	caacgcgaatgaggaagtctggaagttccccaattctgcctgtatatttggaacaaagccac	420
QY	450	tggctattacttcaatgtctcatgatatgtgaaagatgcacaaatcttttcaagaataacct	509
Db	421	tggctattacttcaatgtctcatgatatgtgaaagatgcacaaatcttttcaagaataacct	480
QY	510	gaaacgcgaacgcccgcgcgcgaatgcaccttcgcgaagggcccttcgaagatcacccgaa	569
Db	481	gaaacgcgaacgcccgcgcgcgaatgcaccttcgcgaagggcccttcgaagatcacccgaa	540
QY	570	gaccgcgcgaacagttgcgcagacctgacgcctctgcacaaatgaccttgaagacacgaataaa	629
Db	541	gaccgcgcgaacagttgcgcagacctgacgcctctgcacaaatgaccttgaagacacgaataa	600
QY	630	ggaagatgatcatgtccgcagaaagccgttggaagaaagacgcgccttgatcacagcuaaaaa	689
Db	601	ggaagatgatcatgtccgcagaaagccgttggaagaaagacgcgccttgatcacagcuaaaad	660
QY	690	aagttaacgcgaaggaatctcagccacttgaggttggaagggtctgaaagaaagacgaagcat	749
Db	661	aagttaacgcgaaggaatctcagccacttgaggttggaagggtctgaaagaaagacgaagcat	720
QY	750	gattgatcaagaaatctgattgagacgctccagatgaaaacctttgacatctacttctccaat	809
Db	721	gattgatcaagaaatctgattgagacgctccagatgaaaacctttgacatctacttctccaat	780
QY	810	caagaatttccgcgtgccacaggggtctttagcaatgtgctatgaatttgcacaaatctctga	869
Db	781	caagaatttccgcgtgccacaggggtctttagcaatgtgctatgaatttgcacaaatctctga	840
QY	870	ggtcccttccaaagaaagaaatctgtccaaatggagcccaagtctgaaaaaatctctatcttt	929
Db	841	ggtcccttccaaagaaagaaatctgtccaaatggagcccaagtctgaaaaaatctctatcttt	900
QY	930	gaaatctctctgcgcgtctgcggggggaatgagcaatcttgcgaacttcaaacctccac	989
Db	901	gaaatctctctctgcgcgtctgcggggggaatgagcaatcttgcgaacttcaaacctccac	960
QY	990	cgacacatgacgaaagaaatcttctccctgtctgcgtccacacatgcttgaactgttcaacta	1049
Db	961	cgacacatgacgaaagaaatcttctccctgtctgcgtccacacatgcttgaactgttcaacta	1020
QY	1050	catgttccaaagcatcatcagctttgtccaaagttatctctacttcaagacattgacctat	1109
Db	1021	catgttccaaagcatcatcagctttgtccaaagttatctctacttcaagacattgacctat	1080
QY	1110	cgaaagacacgaatctctccctgcgttgaagggggcccttctgaagctgttgcacatgaaatcaa	1169
Db	1081	cgaaagacacgaatctctccctgcgttgaagggggcccttctgaagctgttgcacatgaaatcaa	1140
QY	1170	cacacatgttccagcggaagacatgaaactggaagatgttgcgcgcgcgttcttactgttga	1229
Db	1141	cacacatgttccagcggaagacatgaaactggaagatgttgcgcgcgcgttcttactgttga	1200
QY	1230	agacacatgaaatgtgcttccagcaacttcttacttgaagcccatgtatgaattctccactat	1289

Db 472 aagtcgagagcgcacaaatctgccttgatctgagagacagccactgctatcaattca 531
 QY 464 atgtatgacatgtgaaatgtcgaagagctcttccagagagccatgtgaacgcaagccc 523
 Db 522 atgtatgacatgtgaaatgtcgaagagctcttccagagagccatgtgaacgcaagccc 591
 QY 524 gcttaagttgcccctccgagagagccttgagagatcaccggaagagcccgagagagt 583
 Db 592 gcttgaagttgcccctccgagagagccttgagagatcaccggaagagcccgagagagt 651
 QY 564 gccagagctgcgcctgcgcgaagtgactctgagagagcagatgaagaagagatgtatcagt 643
 Db 652 gccagagctgcgcctgcgcgaagtgactctgagagagcagatgaagaagagatgtatcagt 711
 QY 644 ccgagagagccgctgagagagagagagcgccttgatcacaagcggaagaaagatgaacgagag 703
 Db 712 ccgagagagccgctgagagagagagagcgccttgatcacaagcggaagaaagatgaacgagag 771
 QY 704 ggaactcagccactgagagagtgacagagagcagagagagagagagagagagagagagagc 763
 Db 772 ggaactcagccactgagagagtgacagagagcagagagagagagagagagagagagagagc 831
 QY 764 tgatgagagcgtcagc 823
 Db 832 tgatgagagcgtcagc 891
 QY 824 tgcagc 883
 Db 892 tgcagc 951
 QY 884 aagaagagcgtcagc 943
 Db 952 aagaagagcgtcagc 1011
 QY 944 agctcagc 1003
 Db 1012 agctcagc 1071
 QY 1004 aagaagatcttcccttgctgcgcacatgagagagagagagagagagagagagagagagc 1063
 Db 1072 aagaagatcttcccttgctgcgcacatgagagagagagagagagagagagagagagagc 1131
 QY 1064 tcatagagcttgcacaaatcatctctacttacttacttacttacttacttacttacttactt 1123
 Db 1132 tcatagagcttgcacaaatcatctctacttacttacttacttacttacttacttacttactt 1191
 QY 1124 cccctgctgaagc 1183
 Db 1192 cccctgctgaagc 1251
 QY 1184 cggagc 1243
 Db 1252 cggagc 1311
 QY 1244 gcttccacaaacttacttacttacttacttacttacttacttacttacttacttacttactt 1303
 Db 1312 gcttccacaaacttacttacttacttacttacttacttacttacttacttacttacttactt 1371
 QY 1304 agctcagatgaagc 1363
 Db 1372 agctcagatgaagc 1431
 QY 1364 cagagtgctgagc 1423
 Db 1432 cagagtgctgagc 1491
 QY 1424 agctctatcatatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1483
 Db 1492 agctctatcatatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1551
 QY 1484 tggctatgcttaacagc 1543
 Db 1552 tggctatgcttaacagc 1611

QY 1544 tccagagatatacaccctcttctacagccctctcaagagagagagagagagagagagagagag 1603
 Db 1612 tccagagatatacaccctcttctacagccctctcaagagagagagagagagagagagagagag 1671
 QY 1604 gcttaagc 1663
 Db 1672 gcttaagc 1731
 QY 1664 cccgagc 1723
 Db 1732 cccgagc 1791
 QY 1724 tctctcccttaagc 1783
 Db 1792 tctctcccttaagc 1851
 QY 1784 tgc 1843
 Db 1852 tgc 1911
 QY 1844 gcaactgagc 1903
 Db 1912 gcaactgagc 1971
 QY 1904 cctatgagc 1963
 Db 1972 cctatgagc 2031
 QY 1964 aagc 2023
 Db 2032 aagc 2091
 QY 2024 tgaagcttcttacttacttacttacttacttacttacttacttacttacttacttacttactt 2083
 Db 2092 tgaagcttcttacttacttacttacttacttacttacttacttacttacttacttacttactt 2151
 QY 2084 tcttccagc 2143
 Db 2152 tcttccagc 2211
 QY 2144 tgg 2146
 Db 2212 tgg 2214
 RESULT 4
 AAX56242
 ID AAX56242 standard: cDNA: 2910 BP.
 XX
 AC AAX56242:
 XX
 DT 16-JUL-1999 (first entry)
 XX
 DE Human vitamin D receptor related gamma protein encoding cDNA.
 XX
 KW Human: vitamin D receptor related protein; VDR; obesity; diabetes;
 KW anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;
 KW hypercholesterolaemia; hyperlipoproteinaemia; osteoporosis; tumour;
 KW hyperproliferative skin disorder; hyperthyroidism; ss.
 OS Homo sapiens.
 XX
 PN W09919354-A1.
 XX
 PD 22-APR-1999.
 XX
 PF 31-AUG-1998: 98WO-SE01548.
 XX
 PR 31-MAR-1998: 98SH-0001148.
 PR 14-OCT-1997: 97SE-0003745.
 XX
 PA (PIIA) PHARMACIA & UPJOHN AB.

[illegible]

Db	2070	logqaaagaaatgccctccaaatcccaatcaaatgaatggtccaaaggtatgcgaagaaacacacagacaa	2129
QY	1973	aaqaaagagcaatctgaagatctatagcccaaataccacagttttgttcgttcctaaatcttt	2032
Db	2130	aaagaaagagcaatctgaaggtatagcccaaataccacagttttgttcgttcctaaatcttt	2189
QY	2033	tcattgcaaacctctaaatagtcctgtctcccaattcccaatccgtttccctccctctccgaag	2092
Db	2190	tcattgcaaacctctaaatagtcctgtctcccaattcccaatccgtttccctccctctccgaag	2249
QY	2093	ctgcttctgaagctcccaagcctgtactcattccgaagatgcataagatattctgaag	2146
Db	2250	ctgcttctgaagctcccaagcctgtactcattccgaagatgcataagatattctgaag	2303
RESULT 6			
Db	AAK56243	AAK56243 standard; cDNA; 2802 bp.	
XX	AAK56243:		
XX	16-JUL-1999	(first entry)	
XX	DE	Human vitamin D receptor related gamma 2 protein encoding cDNA.	
XX	XX	Human: vitamin D receptor related protein; VDR; obesity; diabetes;	
XX	XX	anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;	
XX	XX	hypercholesterolaemia; hyperlipoproteinaemia; osteoporosis; tumour;	
XX	XX	hyperproliferative skin disorder; hypothyroidism; ss.	
XX	XX	Homo sapiens.	
XX	XX	W0919354-A1.	
XX	EN	22-APR-1999.	
XX	PD	41-AUG-1998: 98WO-SE01548.	
XX	PF	31-MAR-1998: 98SE-0001148.	
XX	PR	14-OCT-1997: 97SE-0003745.	
XX	XX	(PNA) PHARMACIA & UPJOHN AB.	
XX	XX	betkenstam A. Dahlberg M:	
XX	XX	WPI: 1999-40250R/25.	
XX	XX	P-PSDB: AAY09516.	
XX	XX	New vitamin D receptor related (VDR) polypeptides, useful for	
XX	XX	treating obesity, diabetes, anorexia and rheumatoid arthritis	
XX	XX	claim 2; Page 21-22; 35pp; English.	
XX	XX	The present sequence encodes a human vitamin D receptor related (VDR)	
XX	XX	polypeptide. Human VDR polypeptides and substances which affect VDR	
XX	XX	signal transduction, can be used for treating metabolic, proliferative	
XX	XX	or inflammatory conditions. They can be used in the manufacture of a	
XX	XX	medicament for treating the following conditions: obesity, diabetes,	
XX	XX	anorexia, lipoprotein defects, hyperlipidaemia, hypercholesterolaemia or	
XX	XX	hyperlipoproteinaemia and osteoporosis, rheumatoid arthritis, benign and	
XX	XX	malign tumours, hyperproliferative skin disorders or hypothyroidism.	
XX	XX	Nucleic acid vectors encoding for expression of a VDR polypeptide can	
XX	XX	be used for treating metabolic, proliferative or inflammatory conditions	
XX	XX	by introducing them into a mammal. The introduced nucleic acid is then	
XX	XX	capable of transforming a cell in vivo and then polypeptide is expressed	
XX	XX	A substance affecting VDR signal transduction, such as an agonist or	
XX	XX	antagonist can be used for the manufacture of a medicament for treating	
XX	XX	metabolic, proliferative or inflammatory condition.	
XX	XX	N.B. The specification specifically claims the VDR nucleic acid and	
XX	XX	polypeptide sequences given in figures 1, 4, 7 and 8, but no figures	
XX	XX	are given in the specification.	

Sequence 2802 BP; 723 A; 715 C; 755 G; 609 T; 0 other;

Query Match 87.0%; Score 1866.6; DB 20; Length 2802;

Best Local Similarity 99.5%; Pred. No. 0; Mismatches 9; Indels 0; Gaps 0;

Matches 1872; Conserved

266 acttaccacacagcagctccagagagcccaagacgaacccctgaggtgagaccacgaagaa 325
139 agtctgcttcaaacatccaagagcccaagacgaacccctgaggtgagaccacgaagaa 198
326 gctggaacacagctgactcttgacacttgagacacagagctcgtctccctgaaagccca 385
139 gctggaacacagctgactcttgacacttgagacacagagctcgtctccctgaaagccca 258
386 gttcaacgacagatgagaaatccgaagatcccaaatctccgctgtatctatgagaaagaa 445
259 gttcaacgacagatgagaaatccgaagatcccaaatctccgctgtatctatgagaaagaa 318
446 ccactggtctatcacttcaatgtcatgacatgtgaagatgcaaggtcttttcaagagag 505
319 ccactggtctatcacttcaatgtcatgacatgtgaagatgcaaggtcttttcaagagag 378
506 ccataaagcagacgcccagctgaaggtgcccctccggaagagagccttgcaaatcaccc 565
379 ccataaagcagacgcccagctgaaggtgcccctccggaagagagccttgcaaatcaccc 438
566 ggaagacccgagcagatgacagctccgctccgctcagcaatgctctgagagagagcatga 625
439 ggaagacccgagcagatgacagctccgctccgctcagcaatgctctgagagagagcatga 498
626 aaaaagaaatgactctgctccgagagcctgtgagagagagagagccttgatcaagcgga 685
499 aaaaagaaatgactctgctccgagagcctgtgagagagagagagccttgatcaagcgga 558
686 aaaaagaaatgactctgctccgagagcctgtgagagagagagagccttgatcaagcgga 745
559 aaaaagaaatgactctgctccgagagcctgtgagagagagagagccttgatcaagcgga 618
746 gaaatcatgactcagagatgactcagacatgaaacacclttacactaccltctccc 805
619 gaaatcatgactcagagatgactcagacatgaaacacclttacactaccltctccc 678
806 atttcaagaatttccgctgacagaggtgctttagaagtgagctgacagatttccgaagcttc 865
679 atttcaagaatttccgctgacagaggtgctttagaagtgagctgacagatttccgaagcttc 738
866 tgcaggtcccatcagagagagagagctgacagatgagacagctcagaaagatcctgtct 925
739 tgcaggtcccatcagagagagagagctgacagatgagacagctcagaaagatcctgtct 798
926 ctttgaagatctctctcagcagctcagagagagagatgacagatctcttgaactcaaacccc 985
799 ctttgaagatctctctcagcagctcagagagagagatgacagatctcttgaactcaaacccc 858
986 cagcgcacagctgagcagagagagatctcttccctgtgcccacatgagctgacatgacaa 1045
859 cagcgcacagctgagcagagagagatctcttccctgtgcccacatgagctgacatgacaa 918
1046 cctacatgttcaaaagcagatcagcttgcacaaatcttccatcttcaagacatcttcc 1105
919 cctacatgttcaaaagcagatcagcttgcacaaatcttccatcttcaagacatcttcc 978
1106 ccatcgagagacagatctcctctcagagagagagccttccagctgtatcacaatgagat 1165
979 ccatcgagagacagatctcctctcagagagagagccttccagctgtatcacaatgagat 1038
1166 tcaaacacagctgttcaaacgagagagatgaaacctgagagatgagccgctgtccctactgct 1225
1039 tcaaacacagctgttcaaacgagagagatgaaacctgagagatgagccgctgtccctactgct 1098
1226 tgaagaacacagctgttccacacacattctatgagacacagctgaaattccact 1285

1099 tgaagaacacagctgttccacacacattctatgagacacagctgaaattccact 1158
1286 acatgcttgaagaagctgacagctgacatgagagagagatgagctgagagagagagagag 1445
1159 acatgcttgaagaagctgacagctgacatgagagagagatgagctgagagagagagagag 1218
1346 tctctcccccagaccccccagagctgctgacagagagagagagagagagagagagagag 1405
1219 tctctcccccagaccccccagagctgctgacagagagagagagagagagagagagagag 1278
1406 aatcgacatcttcaaatgctcactatgatatcaatcagcagagagagagagagagagagag 1465
1279 aatcgacatcttcaaatgctcactatgatatcaatcagcagagagagagagagagagagag 1338
1466 tctctcccccagaccccccagagctgctgacagagagagagagagagagagagagagag 1525
1339 tctctcccccagaccccccagagctgctgacagagagagagagagagagagagagagag 1498
1526 cccagagagctgctgacagatcagagagagagagagagagagagagagagagagagagag 1585
1399 cccagagagctgctgacagatcagagagagagagagagagagagagagagagagagagag 1458
1586 tctcagcagatcag 1645
1459 tctcagcagatcag 1518
1646 gacccag 1705
1519 gacccag 1578
1706 gacccag 1765
1579 gacccag 1638
1766 cctcag 1825
1639 cctcag 1698
1826 gacccag 1885
1699 gacccag 1758
1886 ccttctcttcttcaaaag 1945
1759 ccttctcttcttcaaaag 1818
1946 caagctgttgaagaag 2005
1819 caagctgttgaagaag 1878
2006 ccaagctgttgaagaag 2065
1879 ccaagctgttgaagaag 1938
2066 tccacatgttccctctctcagcagctcagagagagagagagagagagagagagagagagag 2125
1939 tccacatgttccctctctcagcagctcagagagagagagagagagagagagagagagagag 1998
2126 cagctgacatgagatctgtg 2146
1999 cagctgacatgagatctgtg 2019

RESUL.T 7
AA78808
ID AAX78808 standard; cDNA: 2850 BP.
XX
AC AAX78808:
XX
DT 06-SEP-1999 (first entry)
XX
DE Human nNR7-1 cDNA.
XX

QY	1766	ccctaaagaagaacttgggtgcaccccccaccccccgaattcaattcgttaaggaatggaagccaca	1825
Db	1663	ccctaaagaagaacacatgggtgcaccccccaccccccgaattcaattcgttaaggaatggaagccaca	1742
QY	1826	gaactctaacgtgagaaatgtaactgaactctgtagtcaagaccatcagaagaagcaatgttc	1885
Db	1743	gattcttaactgtgagaatgtgaactgaactctgtagtcaagaccatcagaagaagcaatgttc	1802
QY	1886	cccttccctcttaaaagcctctatgtctcgggagagaatccctcagatcccaactaaagt	1945
Db	1803	cccttccctcttaaaagcctctctgtgtctcgggagagaatccctcagatcccaactaaagt	1862
QY	1946	caaaqatctaaaaaggaacacaaagcagcaaaagaatagagcactatgaagttctctatgcccacatar	2005
Db	1863	caaaqatctgtgaaggaacacaaagcagcaaaagaatagagcactatgaagttctctatgcccacatar	1922
QY	2006	ccaagttctgttcggcttccttcgaactcttctcatatgaactctaaatggtctgtctcccaact	2065
Db	1923	ccaagttctgttcggcttccttcgaactcttctcatatgaactctaaatggtctgtctcccaact	1982
QY	2066	tcaccaactcgtttccctcctctctcccaagcctgaactcttctgtggtctccaagcctgaactcatcgg	2125
Db	1983	tcaccaactcgtttccctcctctctcccaagcctgaactcttctgtggtctccaagcctgaactcatcgg	2042
QY	2126	caagtgcatatgagatctctctg	2146
Db	2043	caagtgcatatgagatctctctg	2063

RESULT	8
AAx89080	
ID	AAx89080 standard; DNA; 2068 BP.

AC AAX89080;

DT 14-SEP-1999 (first entry)

Human steroid and xenobiotic receptor (SXR) encoding DNA

KM Nuclear receptor: SXR; steroid and xenobiotic receptor: RXR; human;
 KM retinoid X receptor: P450 gene; steroid hormone; steroid metabolism;
 KM phytoestrogen; calcium-channel blocker; steroid toxicity; tuberculosis
 KM breast cancer; osteoporosis; Cushing syndrome; virilism; hirsutism;
 KM polycystic ovarian disease; cancer; colorectal; prostatic; ss.

05 Homo sapiens.

FH	Key	Location/Qualifiers
200		200 1007

	/★tag-	a
ET		
ET		

XY /note= "Xaa= unknown; the start codon is not indicated"

PN W09935246-A1-
XX

FD 302 (Rev. 1-25-60)

XX

XX

XX

XX

DR P-PSDB; AAY21799.

PT	New steroid and x
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

PT their toxicity
XX

PS Claim 10; Fig 1A; 83pp; English.

XX The invention relates to a novel nuclear receptor polypeptide, designated
CC SXR (steroid and xenobiotic receptor). SXR (i) forms a heterodimer with
CC retinoid X receptor (RXR), (ii) binds to a direct or inverted repeat
CC response element motif based on the half-site AGTCA, (iii) activates
CC transcription through response elements present in steroid-inducible p450
CC genes, in response to a wide variety of natural and synthetic steroid
CC hormones and (iv) is prominently expressed in liver and intestine. SXR
CC regulates expression of catabolic enzymes, in response to many different
CC steroids, and thus affects metabolism. SXR is a broad specificity, low-
CC affinity receptor for reducing excessive levels of steroids in the
CC circulation. (Anti)agonists of SXR are used to regulate metabolism of
CC steroids particularly phytoestrogens or calcium-channel blockers, to
CC reduce steroid toxicity in subjects being treated with steroids, e.g. in
CC cases of tuberculosis (treated with rifampin and related compounds),
CC breast cancer (treated with tamoxifen, raloxifen etc.) or osteoporosis
CC (treated with vitamin K), or to slow metabolism of therapeutic steroids.
CC Also, modulating endogenous SXR is used to treat disease, particularly
CC an agonist is used where endogenous steroid levels are excessive (e.g.,
CC Cushing syndrome; virilism and hirsutism in women; polychydrosis/steroid
CC disease) i.e. by a 17- or 21-hydroxylase deficiency; 3 beta-hydroxysteroid
CC dehydrogenase deficiency, or breast, colorectal or prostatic cancer).
CC While antagonists are used where endogenous steroid levels are too low.
CC Cells that express SXR are used to identify compounds likely to be
CC involved in undesirable drug interactions. Antibodies specific for SXR
CC are used in immunohistochemical testing for studying distribution/
CC expression density of SXR, also for diagnosis and therapeutically as
CC antagonist. The present sequence represents the longest SXR cDNA clone
CC encoding the SXR polypeptide.

XX
XX Sequence 2068 bp: 520 A: 541 C: 586 G: 420 T: 1 other:

Sequence 2068 BP; 520 A; 541 C; 586 G; 420 T; 1 other;

Query Match	81.3%	Score 1744.4	DB 20	Length 2068
Best Local Similarity	99.48%	Pred. No. 0		
Matches 181; Conservative	0	Mismatches	7	Indels 3; Gaps 3

QY	1	tgaatatatagatgaagaagcaaatgttctccatctccggaggaactccfnaactatagatga	450
DB	280	tgaatatatagatgaagaagcaaatgttctccatctccggaggaactccfnaactatagatga	339
QY	61	gaacgaagaagaagcaactgccttacttactgaatctcgaactccgaactccgaactc	120
DB	340	gaacgaagaagaagcaactgccttacttactgaatctcgaactccgaactccgaactc	399
QY	121	aaattttatcatgttgagaaagcaaaagaataaagctaaatactctatctctgaacgaagca	180
DB	400	aaattttatcatgttgagaaagcaaaagaataaagctaaatactctatctctgaacgaagca	459
QY	181	cggctccttgaatgaacgaactccttgaacgaactccttgaacgaacttcttcaadaatgaac	240
DB	460	cggctccttgaatgaacgaactccttgaacgaactccttgaacgaacttcttcaadaatgaac	519
QY	241	cccgaaggagaagatctcgagcaagaagaacttaccacccaagcaatccaaagaagcccaaaaga	300
DB	520	cccgaaggagaagatctcgagcaagaagaacttaccacccaagcaatccaaagaagcccaaaaga	579
QY	301	aacctgagagtgaaagccccaataaagaacgagaaacatgtgttgactttgacatgaaagaac	360
DB	580	aacctgagagtgaaagccccaataaagaacgagaaacatgtgttgactttgacatgaaagaac	639
QY	361	acaaagatctatctctgagaagaagccaaagtgtcaacgcaaatgaagaatccaaagatcccaaa	420
DB	640	acaaagatctatctctgagaagaagccaaagtgtcaacgcaaatgaagaatccaaagatcccaaa	699
QY	421	atcttccctgaatgaggaagaagccacccctgcatatcatcttcaatgatalaagaatga	480
DB	700	atcttccctgaatgaggaagaagccacccctgcatatcatcttcaatgatalaagaatga	759
QY	481	gaatgcaaaagatcttttcagaaagccatgaacgaagcccaagcccaatgaagatcccttc	540
DB	760	gaatgcaaaagatcttttcagaaagccatgaacgaagcccaagcccaatgaagatcccttc	819

[illegible]

D6	1898	1999T GAGACCTTCGAGAGGAGAGCAGCCATCGAAGCCGAGATCCCAGGCC	1957
UY	1680	AAGAACAATGAACATGCACCAAGCCAGCAATGGCCGTGCTGGCTGTCTCCCTAAAGATT	: 739
D6	1958	GAGAACAGATGAGCACTGCCAAGAGCCGCAATGGCCGTGGTGGCTGTCTCCCTAAAGATT	2017
UY	1740	TCTCATCTGAGAGCTGGCTAGCAATCTCTCAAGAGAGCAATGGTGCCCTC	1760
D6	2018	TCTCATCTGAGAGCTGGCTAGCAATCTCTCAAGAGAGCAATGGTGCCCTC	2068
 RESULT 9 AAK59967			
ID	AAK59967	standard; DNA; 1374 BP.	
XX	AAK59967:		
AC	AAK59967:		
XX	04-AUG-1999	(first entry)	
DT			
XX			
DE		DNA encoding an intranuclear receptor protein.	
XX			
KW		Human; intranuclear receptor protein; drug development; diagnosis;	
KN		treatment; SS.	
XX			
OS		Homo sapiens.	
XX			
FN	JP11127872-A.		
XX	18-MAY-1999.		
PD			
XX	07-AUG-1998:	98JP-0224172.	
PY			
XX	11 AUG-1997:	97JP-0240335.	
PK			
XX			
PA	(NLSH) JAPAN TOBACCO INC.		
XX	WPt: 1999-350349/30.		
DK	P-PStDB: AAK15932.		
XX			
PT	New intranuclear receptor protein - useful for drug development and		
DI	diagnosis and treatment of disease		
XX			
PS	claim 4: Page 23 25; 48pp; Japanese.		
XX			
CC	The present sequence encodes a human intranuclear receptor protein.		
CC	The nucleic acid sequence was isolated from a human adult cDNA		
CC	library using a swedish ANO24 derived probe. The protein can		
CC	be used for the development of drugs and diagnostics and treatment		
CC	of various diseases.		
XX			
SQ	Sequence 1474 BP: 329 A; 375 C; 406 G; 264 T; 0 other:		
 Query Match: 64.0%; Score 1374; BH 20; Length 1474; Best Local Similarity 100.0%; Pred. No. 0; Matches 1374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
UY	235	atgaaccccaagggaggaaattcgaagcacagaagacttaccacccaaggaattccaagaaagccca	294
D6	1	gtttggttt	
UY	295	gaagacaacctgaagaaatagacaccaagaagaacttagaacatgccttgttaccatgt	354
D6	61	gaagacaacctgaagaaatagacaccaagaagaagaacttagaacatgccttgttaccatgt	120
UY	355	gaagacaacctgtgttcttgaaagaccgaagtataacacgaatgaagaaagtccgaagt	414
D6	121	gaagacaacctgtgttcttgaaagaccgaagtataacacgaatgaagaaagtccgaagt	180
UY	415	cctcaaatctcgccgtgatattgagacaaaagcccatgccatgccatgccatgccatgccatgccat	474
D6	181	cctcaaatctcgccgtgatattgagacaaaagcccatgccatgccatgccatgccatgccatgccat	240

QY	475	ctgtbaagatgcaagagggctcttttcaggaaggtccatgaaacgnaagccgcgtgtaggtgc	534
Dd	241	ctgtbaagatgcaagagggctcttttcaggaagggccatgaaacgnaagccgcgtgtaggtgc	300
QY	535	cccttcgcgaagcgccgtcttcagagatccaccgcgaagaccgcgcgaatgtccagaccgc	594
Dd	301	cccttcgcgaagcgccgtcttcagagatccaccgcgaagaccgcgcgaatgtccagaccgc	360
QY	595	cgccttcgcgaaggtgccttcggaagcgagatgaaagatgatatcatgtccgcacaagcc	654
Dd	361	cgccttcgcgaaggtgccttcggaagcgagatgaaagatgatatcatgtccgcacaagcc	420
QY	655	gtgtagagagaaagcgccgtcttcatcaagcggaagaaagatgaaacgagaggaatcagcca	714
Dd	421	gtgtagagagaaagcgccgtcttcatcaagcggaagaaagatgaaacgagaggaatcagcca	480
QY	715	ctggagatgacgagagcgctgaaagaaagcgagatgatatgataaggaactatgagacct	774
Dd	481	ctggagatgacgagagcgctgaaagaaagcgagatgatatgataaggaactatgagacct	540
QY	775	cagatgaaacactttgacatctactctcccatcttcaagaatcttcgystgcgcagagtg	834
Dd	541	cagatgaaacactttgacatctactctcccatcttcaagaatcttcgystgcgcagagtg	600
QY	835	cttcagcaatgagctgaggaattgcacagagctctgcgaagcccatcgaggaagaaagctgc	894
Dd	601	cttcagcaatgagctgaggaattgcacagagctctgcgaagcccatcgaggaagaaagctgc	660
QY	895	aagtcgagacgagagtcgcggaagaaagatctgtactctttaaagctctcttcaactgcgggg	954
Dd	661	aagtcgagacgagagtcgcggaagaaagatctgtactctttaaagctctcttcaactgcgggg	720
QY	955	gaagatgcagatgctcttgaaactaaacaccccgccgcgcaagctgcgcggaagaaagatcttc	1014
Dd	721	gaagatgcagatgctcttgaaactaaacaccccgccgcgcaagctgcgcggaagaaagatcttc	780
QY	1015	tcctctgcctcccaatgagctgatactgaactacatagtctcaagagcatcaacagctt	1074
Dd	781	tcctctgcctcccaatgagctgatactgaactacatagtctcaagagcatcaacagctt	840
QY	1075	gccaaagcatctctactctcaaggaattgcgcatactgaagacaaagatctccctgtctgaag	1134
Dd	841	gccaaagcatctctactctcaaggaattgcgcatactgaagacaaagatctccctgtctgaag	900
QY	1135	gagggccgcttcctcgagcgtgtctcaactgaaagatcttcaacacagatcttcaacgcgagagctga	1194
Dd	901	gagggccgcttcctcgagcgtgtctcaactgaaagatcttcaacacagatcttcaacgcgagagctga	960
QY	1195	accttgagagatgtgacgcgagctgtccctactgctcttgaaagacactgcagatgtctccagaa	1254
Dd	961	accttgagagagatgtgacgcgagctgtccctactgctcttgaaagacactgcagatgtctccagaa	1020
QY	1255	cttctactcgagagcccatgagctgaaatctccatctatagctgaaagagctgcagatgtgtag	1314
Dd	1021	cttctactcgagagcccatgagctgaaatctccatctatagctgaaagagctgcagatgtgtag	1080
QY	1315	gagagagatgagctgagatgcagagccatctccctctctcccccagagcccgagatgtgtg	1374
Dd	1081	gagagagatgagctgagatgcagagccatctccctctctcccccagagcccgagatgtgtg	1140
QY	1375	cagcaacgcgtgagtgaacacgagctgcagaggaatctgcacattactctaaagctcatcat	1434
Dd	1141	cagcaacgcgtgagtgaacacgagctgcagaggaatctgcacattactctaaagctcatcat	1200
QY	1435	gaatgacaatcgagcccgagcctgctcctaatagtctctgtctccgaaagatataragtatgctc	1494
Dd	1201	gaatgacaatcgagcccgagcctgctcctaatagtctctgtctccgaaagatataragtatgctc	1260
QY	1495	accagagctccgacatctaatgtatgatacacaacccagcgctgctgtgcgataccagaacata	1554
Dd	1261	accagagctccgagcatcaatgtctcagacaacccagcgagctgtgctgtcatalccagaacata	1320

[illegible]

Db	421	caagggcttgagacaagagacagcagatgatcatcaaggagcttgagacgctcaagtggaa	480
Qy	784	acccttggacacatacctctctccattccaaagaatttcgcgtctccacagatattgaacat	843
Db	481	acccttgaacatacctctccattccaaagaatttcgcgtctccacagatattgaacat	540
Qy	844	ggtctcgaagtgcagagatctctgcagaccccatcgagggaggaagatctccaaagtggac	903
Db	541	ggtctcgaagtgcagagatctctgcagaccccatcgagggaggaagatctccaaagtggac	600
Qy	904	caagctccggaagaatctctctcttgaaggtctctctgcagctcgcgagggaggaatgac	963
Db	601	caagctccggaagaatctctctcttgaaggtctctctgcagctcgcgagggaggaatgac	660
Qy	964	agttctcggaaactacaaaccccgacgcgcagatggcggaagaagaattctcccttgctg	1023
Db	661	agttctcggaaactacaaaccccgacgcgcagatggcggaagaagaattctcccttgctg	720
Qy	1024	ccccacatgagctgacatgattcaactcatttcaaaagatattatagcttttccaaatc	1083
Db	721	ccccacatgagctgacatgattcaactcatttcaaaagatattatagcttttccaaatc	780
Qy	1084	atctctcaacttcagggagctgtggccatcgagagacaaattctctctgtctgaaggggccct	1143
Db	781	atctctcaacttcagggagctgtggccatcgagagacaaattctctctgtctgaaggggccct	840
Qy	1144	ttcgaagctgtgtcaacacgaaattcaacacagtttctcaacgcggaagatcgaaactctggag	1203
Db	841	ttcgaagctgtgtcaacacgaaattcaacacagtttctcaacgcggaagatcgaaactctggag	900
Qy	1204	ttgtggccgagctgtctctactgtcttggaagacactgcagctgagcttccagaaacttactg	1263
Db	901	ttgtggccgagctgtctctactgtcttggaagacactgcagctgagcttccagaaacttactg	960
Qy	1264	gagcccatctgcgaaattccactacatgctgaagaagatctcaagctgcatgagagagatat	1323
Db	961	gagcccatctgcgaaattccactacatgctggaagaagatctcaagctgcatgagagagatat	1020
Qy	1324	gtgtgtaattgcagagccatctccctctctctcccgagaccgcgcaggtgtgtctgacacgcg	1383
Db	1021	gtgtgtaattgcagagccatctccctctctctcccgagaccgcgcaggtgtgtctgacacgcg	1080
Qy	1384	gtgtgtgagaccagctgcagaggaacatttcgcattactctgaagtctctactatthaatgcaat	1443
Db	1081	gtgtgtgagaccagctgcagaggaacatttcgcattactctgaagtctctactatthaatgcaat	1140
Qy	1444	cgagccccaagctgtctctgaagtctgttcttccttaaatatgctatgtctccagagctc	1503
Db	1141	cgagccccaagctgtctctgaagtctgttcttccttaaatatgctatgtctccagagctc	1200
Qy	1504	cgagcagcatcaatgctctcagacaccccgagcgtgtgcgcatccagacatataccctctt	1563
Db	1201	cgagcagcatcaatgctctcagacaccccgagcgtgtgtgcgcatccagacatataccctctt	1260
Qy	1564	gctatgcgccctcatgcagaggtgtgttcggatcatcacagatgacgtga	1608
Db	1261	gctatgcgccctcatgcagaggtgtgttcggatcatcacagatgacgtga	1305
RESULT 12			
ID	AAK59972		
AAK59972 standard; DNA: 644 BP.			
XX	AAK59972:		
AC	04-AUG-1999 (first entry)		
DT	SEQ ID 12 of JP1127872.		
XX			
DE			
XX			
KM	Human; intranuclear receptor protein; drug development; diagnosis;		
KM	treatment; ss.		
XX			

[illegible]

Thu Jul 12 14:23:28 2001

us-09-276-935b-13.rng

Page 19

2

117: qb_cst48:*
 118: qb_cst49:*
 119: qb_cst50:*
 120: qb_cst51:*
 121: qb_cst52:*
 122: qb_cst53:*
 123: qb_cst54:*
 124: qb_cst55:*
 125: qb_cst56:*
 126: qb_cst57:*
 127: qb_cst58:*
 128: qb_cst59:*
 129: qb_cst60:*
 130: qb_cst61:*
 131: qb_cst62:*
 132: qb_cst63:*
 133: qb_cst64:*
 134: qb_cst65:*
 135: qb_cst66:*
 136: qb_cst67:*
 137: qb_cst68:*
 138: qb_cst69:*
 139: qb_cst70:*
 140: qb_cst71:*
 141: qb_cst72:*
 142: qb_cst73:*
 143: qb_cst74:*
 144: qb_cst75:*
 145: qb_cst76:*
 146: qb_cst77:*
 147: qb_cst78:*
 148: qb_cst79:*
 149: qb_cst80:*
 150: qb_cst81:*
 151: qb_cst82:*
 152: qb_cst83:*
 153: qb_cst84:*
 154: qb_cst85:*
 155: qb_cst86:*
 156: qb_cst87:*
 157: qb_cst88:*
 158: qb_cst89:*
 159: qb_cst90:*
 160: qb_cst91:*
 161: qb_cst92:*
 162: qb_cst93:*
 163: qb_cst94:*
 164: qb_cst95:*
 165: qb_cst96:*
 166: qb_cst97:*
 167: qb_cst98:*
 168: qb_cst99:*
 169: qb_cst100:*
 170: qb_cst101:*
 171: qb_cst102:*
 172: qb_cst103:*
 173: qb_cst104:*
 174: qb_cst105:*
 175: qb_cst106:*
 176: qb_cst107:*
 177: qb_cst108:*
 178: qb_cst109:*

190: qb_cst110:*
 191: qb_cst111:*
 192: qb_cst112:*
 193: qb_cst113:*
 194: qb_cst114:*
 195: qb_cst115:*
 196: qb_cst116:*
 197: qb_cst117:*
 198: qb_cst118:*
 199: qb_cst119:*
 200: qb_cst120:*
 201: qb_cst121:*
 202: qb_cst122:*
 203: qb_cst123:*
 204: qb_cst124:*
 205: qb_cst125:*
 206: qb_cst126:*
 207: qb_cst127:*
 208: qb_cst128:*
 209: qb_cst129:*
 210: qb_cst130:*
 211: qb_cst131:*
 212: qb_cst132:*
 213: qb_cst133:*
 214: qb_cst134:*
 215: qb_cst135:*
 216: qb_cst136:*
 217: qb_cst137:*
 218: qb_cst138:*
 219: qb_cst139:*
 220: qb_cst140:*
 221: qb_cst141:*
 222: qb_cst142:*
 223: qb_cst143:*
 224: qb_cst144:*
 225: qb_cst145:*
 226: qb_cst146:*
 227: qb_cst147:*
 228: qb_cst148:*
 229: qb_cst149:*
 230: qb_cst150:*
 231: qb_cst151:*
 232: qb_cst152:*
 233: qb_cst153:*
 234: qb_cst154:*
 235: qb_cst155:*
 236: qb_cst156:*
 237: qb_cst157:*
 238: qb_cst158:*
 239: qb_cst159:*
 240: qb_cst160:*
 241: qb_cst161:*
 242: qb_cst162:*
 243: qb_cst163:*
 244: qb_cst164:*
 245: qb_cst165:*
 246: qb_cst166:*
 247: qb_cst167:*
 248: qb_cst168:*
 249: qb_cst169:*
 250: qb_cst170:*
 251: qb_cst171:*
 252: qb_cst172:*
 253: qb_cst173:*
 254: qb_cst174:*
 255: qb_cst175:*
 256: qb_cst176:*
 257: qb_cst177:*
 258: qb_cst178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1006.8	46.9	2525	192	AK018630	AK018630 Mus muscu
2	370.8	17.3	794	24	A1746915	A1746915 u109a10.y
3	343	16.0	343	17	A1248626	A1248626 qh77D12.x
4	325.4	15.2	677	107	AK011108	AK011108 Mus muscu
5	288.8	13.5	300	107	A0099521	A0099521 A0099521
6	222	10.3	638	32	AV651714	AV651714 AV651714
7	214.8	10.0	454	187	N94132	N94132 za25906.r1
8	201	9.4	472	4	AA277370	AA277370 vab1a12.r
9	199.6	9.3	493	121	AA677811	AA677811 da94c06.y
10	185.4	8.6	200	10	AA679591	AA679591 z149d12.s
11	177.6	8.3	837	108	A0140214	A0140214 A0140214
12	166.4	7.8	463	117	AM511148	AM511148 hdd2a05.x
13	165.4	7.7	215	25	AV108557	AV108557 AV108557
14	159.6	7.4	677	169	BF788461	BF788461 602114213
15	158	7.4	689	150	BF577692	BF577692 602092577
16	156.8	7.3	432	112	AM158294	AM158294 za39e07.x
17	154.4	7.2	601	111	AM107536	AM107536 u191a06.y
18	147	6.8	776	168	BF686868	BF686868 602102822
19	143.8	6.7	560	24	A1768052	A1768052 w146b06.x
20	141.6	6.6	654	144	BF102347	BF102347 601751948
21	140.8	6.6	1053	137	BE573696	BE573696 601333441
22	140.4	6.5	806	248	A2716492	A2716492 RPO1-24-1
23	132.8	6.2	540	110	AM012320	AM012320 um07d07.y
24	132	6.2	463	15	A1049299	A1049299 uc85h08.y
25	124.6	5.8	832	155	BC565979	BC565979 602583159
26	118	5.5	724	15	A1046425	A1046425 u065f04.x
27	116.2	5.4	391	250	A2816919	A2816919 2M0085020
28	108.4	5.1	468	6	AA396982	AA396982 mx6a06.r
29	103.6	4.8	950	106	AL559996	AL559996 AL559996
30	103.4	4.8	778	148	BF383394	BF383394 602045056
31	102	4.8	463	2	AA107961	AA107961 mp04d08.r
32	101.6	4.7	859	105	AL522278	AL522278 AL522278
33	101.2	4.7	506	10	AA679391	AA679391 z129a02.s
34	100.8	4.7	616	17	A1238397	A1238397 gh14435.5
35	100.8	4.7	859	106	AL583005	AL583005 AL583005
36	99.6	4.6	683	169	BF783185	BF783185 602109385
37	99.2	4.6	732	104	AJ343915	AJ343915 AJ343915
38	99	4.6	370	15	A1016672	A1016672 ovg6004.x
39	99	4.6	507	4	AA264990	AA264990 LD08686.5
40	98.2	4.6	876	175	BG281373	BG281373 602401868
41	97.6	4.5	690	165	BE266478	BE266478 601192748
42	97.6	4.5	1105	167	BE410116	BE410116 601302405
43	97.4	4.5	560	144	BF080344	BF080344 231159.MA
44	97.2	4.5	589	165	BE275249	BE275249 601122072
45	96	4.5	726	108	A0136157	A0136157 A0136157

ALIGNMENTS

JOURNAL REFERENCE	JOURNAL REFERENCE
2 (sites)	Methods Enzymol. 303, 19-44 (1999)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
Genome Res. 10 (10), 1617-1630 (2000)	
JOURNAL REFERENCE	JOURNAL REFERENCE
3 (sites)	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, F., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Chara, E., Motani, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Okazaki, Y., Muramatsu, M., Inoue, Y., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipipillary sequencer	
Genome Res. 10 (11), 1757-1771 (2000)	
JOURNAL REFERENCE	JOURNAL REFERENCE
4 (sites)	The RIKEN Genome Exploration Research Group Phase II team and FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection	
Nature 409, 685-690 (2001)	
JOURNAL REFERENCE	JOURNAL REFERENCE
5 (phases 1 to 2525)	Submitted (10-Jul-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
Please visit our web site (http://genome-gsc.riken.go.jp/) for further details.	
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGACAGACATCCGACGATCTTTTATTAATTATTCCTCCGCCCC 3'. cDNA was prepared by using triphosphate thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to rot -10.0 and subtraction to rot -185.2. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGACAGACATCCGACGATCTTTTATTAATTATTCCTCCGCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vectors: a modified plasmid pUC19 after bulk excision from lambda phi C1 cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.	
Location/Qualifiers	
1. 2525	
/organism="Mus musculus"	
/strain="C57BL/6J"	
/db_xref="taxon:10090"	
/db_xref="MGI:1337040"	
/db_xref="MGI:189757"	
/clone="9130016118"	
/sex="male"	
/tissue_type="cecum"	
/clone_lib="RIKEN full-length enriched mouse cDNA library"	

FEATURES

source

QY	563	cctggaaagacccgqgaacaagcgcagcgtcgcgccttcgcgaagcgctcgaagaccgca	622
Dd	555	ccccgaatacacacacgaggcacttcccacgccctgcccgtttccgaagtgacctgacga	614
QY	623	tgaagaagagatgatcatqtccgaagcagcgcgtgtaagaagaaqgcgcgttatcaa	681
Dd	615	tcgaacaaacagatgatatgtccgatccgctgtgcacacagagcgccgcttgatcaag	673
RESULT	5		
LOCUS	AD099521		
DEFINITION	AD099521	300 bp mRNA EST 05-APR-2001	
ACCESSION	AD099521	Sugano Homo sapiens cDNA library Homo sapiens cDNA clone HS107557 similar to Homo sapiens orphan nuclear receptor (PARI)	
VERSION	AD099521.1	GI:13550650	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Suzuki,Y., Tsunoda,T., Taira,H., Miushima-Sugano,J., Sese,J., Hata ,H., Ota,T., Iscogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo ,K., Suyama,A. and Sugano,S. In silico mapping of the 5'-ends of human mRNAs using full-length enriched and 5'-end enriched cDNA libraries constructed by oligo-capping method Unpublished (2001) Contact: Yutaka Suzuki Department of Virology Institute of Medical Science, University of Tokyo 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan Email: yusuzuki@ims.u-tokyo.ac.jp Suzuki.Y., Yoshitomo-Nakayawa,K., Maruyama,K., Suyama,A. and Sugano,S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).		
FEATURES	source	1..300 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="HS107557" /cdone_lib="Sugano Homo sapiens cDNA library"	
BASE COUNT	91 a 73 c 76 g 60 t		
ORIGIN			
Query Match	13.5%	Score 288.8;	DB: 107; Length 300;
Best Local Similarity	97.7%;	Pred. No. 4,1e-64;	
Matches 293;	Conservative 0;	Mismatches 7;	Indels 0;
Gaps	0;		
QY	20	aagattgctcraataccgaggaatcataactatgaactaagaacgaagaaqaacac	79
Dd	1	AGATTGTCTCATATCCGGGCCAATCAACTATGACTAGGAGCGGAAGAGACGACT	60
QY	80	gccttaactcaagtgaalctcgagcctcagcgcgcgaacccaagtgttccacgtaga	139
Dd	61	GCCTTACTTCAGTGGGATCTCGGCTCAGCCTCAACGCCAACGTTTCACAGTGAGAA	120
QY	140	agcagaagaatagaactaatcctcgcctcgaacgaagcagcgcgtcctcttlaagcla	199
Dd	121	AGCAAGAGNATAGCTAATACTCTCTGTCTGAAAAAOCGACGCGTTCTTGSTAAA	180
QY	200	cctcttgatcgaactcttgaacccgagatgttcaaaatgaaccccccaggaagatcgag	259
Dd	181	CTCCTTGATCAGATCCTTTSCACCGGATGTGTTCAAGTGGACCCTCAGGAGAA	240
QY	260	caaaaacttacccaccaacacgactcaagaagcgcacgaagcgaacttgaatgaaccca	319
Dd	241	CAAAGACTTACCCCAACGAGTTCGAAGAAGTAGCAACAACCAAACTGAGAGTGACACCA	300

RESULT	6	AV651714	LOCUS	AV651714	648 bp	mRNA	EST	07-SEP-2000
DEFINITION		AV651714	GLC Homo sapiens cDNA clone GLCSC03 3', mRNA sequence.					
ACCESSION		AV651714						
VERSION		AV651714.1	GI:9872728					
KEYWORDS		EST.						
SOURCE		human.						
ORGANISM		Homo sapiens						
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
AUTHORS		Qian,B., Wu,T., Huang,Q., Huang,C., Kan,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Yu,D., Zeng,D., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.						
TITLE		Homo sapiens cDNA clone						
JOURNAL		Unpublished (2000)						
COMMENT		Contact: Zengqiang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhaogjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzq@chgc.sh.cn This clone is available at CHGC in Shanghai.						
FEATURES		source						
		1..638						
		/organism="Homo sapiens"						
		/db_xref="taxon:9606"						
		/clone="GLCSC03"						
		/clone_lib="GLC"						
		/rissue_type="corresponding non cancerous liver tissue"						
		/dev_stage="Adult"						
		/lab_host="SOLR"						
		/note="Vector: pBluescript sk(-); Site_1: EORI; Site_2: XhoI"						
BASE COUNT		160 a 149 c 162 g 167 t						
ORIGIN								
Query Match		10.3%: Score 222; DH 32; Length 638;						
Best Local Similarity		100.0%: Pred.No. 1;le 46;						
Matches 222: Conservative		0: Mismatches 0: Indels 0: Gaps 0						
QY	1925	ectcgatccctaaagtcaagtgatgaagagcaagcagcaagatgaagcat	1984					
DB	1	CCTCATGTCCTAAAGTGTCAGGTGTGGAGGAGGACCAAGCAGCATAGCCAT	60					
QY	1985	ctcgaagctcagccacataccacagtttctgcctctcagatcttcattgctact	2044					
DB	61	CTGGGCTATAGCCACATACCCAGCTTGCTGCTCTCATGATCTTTCATGTACCG	120					
QY	2045	caataagctcgtctcccaacttcccaactgttccctctcttcgaagtgatctatgag	2104					
DB	121	CTAATAGTCTGTCTCCCACTTCCCACTGTGCTCCCTCTCTCCAGGCTGTGTGGG	180					
QY	2105	ctccagagctataactcagcagatgagatgatatatgag	2146					
DB	181	CTCCAGGCTGTACTCATGCGCAGCATGATGATCTGTGG	222					
RESULT	7							
LOCUS		N94132	454 bp	mRNA	EST		05-APR-1996	
DEFINITION		zsf25g06.r1 Soares fetal liver spleen INELIS Homo sapiens cDNA clone						
ACCESSION		N94132						
VERSION		N94132.1	GI:1266441					
KEYWORDS		EST.						
SOURCE		human.						
ORGANISM		Homo sapiens						
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						

Thu Jul 12 14:23:32 2001

us-09-276-935b-13.rst

Page 13

4

GenCore Version 4.5
Copyright (c) 1993-2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2001, 12:32:45 ; Search time 104.84 Seconds
(without alignments)
3792.147 Million cell updates/sec

Title: US-09-276-935B-13

Perfect score: 2146
Sequence: 1 tgaatataagtgagagaca.....aaqtqatagatctctgtg 2146

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736.6	34.3	801	4	US-09-276-531-111 Sequence 111, App
2	278.8	13.0	2043	1	US-07-737-736B-6 Sequence 6, Appli
3	272.8	12.7	1399	1	US-07-737-736B-5 Sequence 5, Appli
4	170.6	7.9	1450	1	US-08-459-489-1 Sequence 1, Appli
5	170.6	7.9	1450	1	US-08-458-686-1 Sequence 1, Appli
6	170.6	7.9	1450	1	US-07-843-350C-1 Sequence 1, Appli
7	170.6	7.9	1450	1	US-08-330-283-1 Sequence 1, Appli
8	97.6	4.5	1898	1	US-08-342-411A-1 Sequence 1, Appli
9	97.6	4.5	1979	2	US-08-649-619B-1 Sequence 1, Appli
10	97.6	4.5	2030	1	US-08-330-283-1 Sequence 1, Appli
11	97.6	4.5	2030	1	US-08-330-283-1 Sequence 1, Appli
12	97.6	4.5	2030	2	US-08-646-248-1 Sequence 1, Appli
13	97.6	4.5	2030	5	PCT-US95-13924-1 Sequence 1, Appli
14	97.6	4.5	2030	5	PCT-US95-13931-1 Sequence 1, Appli
15	97.2	4.5	1813	5	PCT-US94-12883-1 Sequence 3, Appli
16	91.4	4.3	2928	5	US-08-095-728B-1 Sequence 3, Appli
17	91.4	4.3	2928	5	PCT-US92-02320A-3 Sequence 3, Appli
18	91.4	4.3	2940	6	Patent No. 5171671
19	91.4	4.3	3036	1	US-08-306-691B-52 Sequence 52, Appli
20	91.4	4.3	3036	2	US-08-095-728B-1 Sequence 1, Appli
21	91.4	4.3	3036	5	PCT-US92-02320A-1 Sequence 1, Appli
22	91.4	4.3	3511	3	US-08-892-747-13 Sequence 13, Appli
23	89.8	4.2	704	2	US-08-592-381-5 Sequence 5, Appli
24	89.8	4.2	2940	2	US-08-592-381-5 Sequence 5, Appli
25	85.2	4.0	1959	1	US-08-342-411A-3 Sequence 3, Appli
26	85.2	4.0	1959	5	PCT-US94-12883-4 Sequence 4, Appli
27	83.8	3.9	2658	2	US-08-592-383-3 Sequence 3, Appli

28	80.8	3.8	1860	2	US-08-372-652-7 Sequence 7, Appli
29	80.8	3.8	1860	5	PCT-US95-16311-7 Sequence 7, Appli
30	79.6	3.7	1659	1	US-08-333-358-7 Sequence 7, Appli
31	79.6	3.7	1659	1	US-08-463-694-7 Sequence 7, Appli
32	79.6	3.7	1659	1	US-08-694-501-7 Sequence 1, Appli
33	77.6	3.6	2970	5	PCT-US92-06391-1 Sequence 1, Appli
34	77.6	3.6	1688	2	US-08-649-619B-2 Sequence 2, Appli
35	74.8	3.5	1649	2	US-08-466-120-1 Sequence 1, Appli
36	74.8	3.5	1649	5	PCT-US94-07266-1 Sequence 1, Appli
37	74.4	3.5	1576	6	5260432-1 Patent No. 5260432
38	74.4	3.5	2241	4	US-09-144-759-17 Sequence 17, Appli
39	74.4	3.5	2295	4	US-09-144-759-19 Sequence 19, Appli
40	74.4	3.5	2301	4	US-09-144-759-21 Sequence 21, Appli
41	73.2	3.4	2468	1	US-08-333-358-11 Sequence 11, Appli
42	73.2	3.4	2468	1	US-08-463-694-11 Sequence 11, Appli
43	73.2	3.4	2468	1	US-08-694-501-11 Sequence 11, Appli
44	72.6	3.4	1893	6	5438126-1 Patent No. 5438126
45	68.6	3.2	1677	2	US-08-372-652-14 Sequence 14, Appli

ALIGNMENTS

RESULT 1
US-09-276-531-111
Sequence 111, Application US/09276531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preethi
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Koopa
APPLICANT: Guebler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: INVOYE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276, 531
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/079, 677
FILING DATE: March 27, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lynn E. Murty, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 801 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SINFET03


```

DB 392 GTACAGCCTAAGAGGAGATCATATGAAGCAAAAGAGAGAGCGCTTGAACACACT 451
QY 715 CTGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774
DB 452 CTGAG---GCCCAACTATCTGACAGCAACAGCACATCATCTCTCTGACGCC 508
QY 775 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 829
DB 509 CACCAACAAACCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
QY 830 -GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888
DB 569 ATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 628
QY 889 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948
DB 629 TCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 688
QY 949 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1008
DB 689 TCCAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 748
QY 1009 ATCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1068
DB 749 CTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 808
QY 1069 AGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128
DB 809 GCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 868
QY 1129 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1188
DB 869 CTTAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 928
QY 1189 ACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239
DB 929 GATATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 988
QY 1240 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1299
DB 989 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1048
QY 1300 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1359
DB 1049 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1108
QY 1360 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1419
DB 1109 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1168
QY 1420 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1479
DB 1169 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1228
QY 1480 ATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1531
DB 1229 ATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1280

```

```

CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/737,736B
FILING DATE: 19910730
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Schwartz, Carl R.
REGISTRATION NUMBER: 29,417
REFERENCE/DOCKET NUMBER: 96-296-2185-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 414-277-5715
TELEFAX: 414-277-5774
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1399 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
AUTHORS: Baker, Andrew R.
AUTHORS: McDonnell, Donald P.
AUTHORS: Hughes, Mark
AUTHORS: Crisp, Tracey M.
AUTHORS: Mangelsdorf, David J.
AUTHORS: Haussler, Mark R.
AUTHORS: Pike, J. W.
AUTHORS: Shine, John
AUTHORS: O'Malley, Bert W.
TITLE: Cloning and expression of full-length cDNA
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 85
PAGES: 3294-3298
DATE: May-1988
US-07-737-736B-5

```

```

Query Match 12.7%; Score 272.8; DH 1; Length 1399;
Best Local Similarity 55.1%; Pred. No. 3,2076;
Matches 638; Conservative 0; Mismatches 487; Indels 44; Gaps 4;

```


[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 4.40)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: DS/08/459,489
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/843,450
FILING DATE: February 26, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/EXCERPT NUMBER: 00786/126001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1450
TYPE: nucleic acid
STRANDNESS: single
TOPLOGY: linear
DS 08-459-489-1

Query Match	7.98	Score 170.67	DB 1	Length 1450
Best Local Similarity	57.58	Prod. No. 7.70-4.2		
Matches 427	Conservative 0	Mismatches 233	Indels 3	Gaps 1
QY 1016	CTCTAGCTACACATCTTCGACACATCAACATCTTCAAGCAATCTGAACTATCAACTTAA	1075		
DB 749	CTCTAGCTACACATCTTCGACACATCAACATCTTCAAGCAATCTGAACTATCAACTTAA	798		
QY 1076	CTCAAGATCATCTCTACTCAAGAACTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1145		
DB 799	CTAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	858		
QY 1116	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1195		
DB 859	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	918		
QY 1146	CTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1252		
DB 919	CTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	978		
QY 1254	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1312		
DB 979	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1038		
QY 1413	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1472		
DB 1049	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1098		
QY 1474	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1432		
DB 1099	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1158		
QY 1484	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1492		
DB 1159	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1218		
QY 1494	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1552		
DB 1219	CAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAGCAATCTTCAAG	1278		

QY 1553 tacacccttgcctacgacccctcatgag 1581
DB 1279 TGTCTCCATGATGCGCTGCTCCAGAG 1307

RESULT 5

US-08-458-686-1

Sequence 1, Application US/08458686
Patent No. 5710017
GENERAL INFORMATION:
APPLICANT: David D. Moore et al.
TITLE OF INVENTION: CAR RECEPTORS AND RELATED
TITLE OF INVENTION: MOLECULES AND METHODS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08458,686
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/843,350
FILING DATE: February 26, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/126001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1450
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-458-686-1

Query Match 7.9%; Score 170.6; DB 1; Length 1450;
Best Local Similarity 57.5%; Pred. No. 7.7e-32;
Matches 327; Conservative 0; Mismatches 239; Indels 3; Gaps 1;

QY 1016 cccgtgtgcccacatggtgaatgtaacctaactgttcaaaaggaatcattcaatttg 1075
DB 739 CTCTGTGTCACACACTTCGACATCAACACTTTCATGCTACTGCAAGTCAATCAAGTTTA 798
QY 1076 ccaaaatcctcctactcagagactccagagagagagagagagagagagagagagag 1135
DB 799 CTAAAGACCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 858
QY 1136 gggcccttccagctgtatcaactgagatcaacacagatgttcaacgagagagagagag 1195
DB 859 GAGGACCTTGGAAATCTGTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 918
QY 1196 ccttgagatgagcagctgtcctcactgcttgagaagacagctgagat---agcttccagc 1252
DB 919 ACTTCTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978
QY 1253 aactctatgagagcctgagaaatctcactatgcttgagagagagagagagagagagag 1312

DB 979 TAGAGTTTTGGAGTTGCTCTTTCACTTCATGCAACACTAGCAAAACTGAGTTCCAG 1078
QY 1313 aagagagatatactatgag 1372
DB 1039 AGCTGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1098
QY 1373 tgcagacccgctgag 1432
DB 1099 CCGAGAGATGAGATTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1198
QY 1433 ttgaatgcaatgag 1492
DB 1159 TCAAGGCGCAGACAG 1218
QY 1493 tcaag 1552
DB 1219 TGGTGAGCTTGGAGCAATATGAGCGGCTGAGCAATGCAATGCAATGCAATGCA 1278
QY 1553 tacacccttgcctacgacccctcatgag 1581
DB 1279 TGTCTCCATGATGCGCTGCTCCAGAG 1307

RESULT 6

US-07-843-350C-1

Sequence 1, Application US/07843350C
Patent No. 5756448
GENERAL INFORMATION:
APPLICANT: David D. Moore et al.
TITLE OF INVENTION: CAR RECEPTORS AND RELATED
TITLE OF INVENTION: MOLECULES AND METHODS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: WordPerfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/843,350C
FILING DATE: February 26, 1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/126001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1450
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-07-843-350C-1

Query Match 7.9%; Score 170.6; DB 1; Length 1450;
Best Local Similarity 57.5%; Pred. No. 7.7e-32;
Matches 327; Conservative 0; Mismatches 239; Indels 3; Gaps 1;

QY 1016 cccgtgtgcccacatggtgaatgtaacctaactgttcaaaaggaatcattcaatttg 1075

[illegible]

```

1 TELEFAX: (617) 542 8906
2 TELFX: 200154
3 INFORMATION FOR SEQ ID NO: 1:
4 SEQUENCE CHARACTERISTICS:
5 LENGTH: 1450
6 TYPE: NUCLEIC ACID
7 STRANDEDNESS: SINGLE
8 TOPLOGY: Linear
9 DEFINITION:
10
11 Query Match: 7.9%; Score 170.6; DB 5; Length 1450;
12 Best Local Similarity 57.5%; Prod No. 7.7e-32;
13 Matches 327; Conservative 0; Mismatches 239; Indels 3; Gaps
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1
```


COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/342,411A
 FILING DATE: 18-NOV-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: KITCHELL, BARBARA S.
 REGISTRATION NUMBER: 33,928
 REFERENCE/DOCKET NUMBER: ARCD154
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (713) 789-2679
 TELEX: 79-0924
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1898 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 71..1450
 US-08-342-411A-1

Query Match 4.5%; Score 97.6; DB 1; Length 1898;
 Best Local Similarity 47.6%; Pred. No. 2.2e-14;
 Matches 595; Conservative 0; Mismatches 569; Indels 76; Gaps 7;

412 gctcccaaatctgcctgctatctgaggaagagccactgctatcaatgtaatg 471
 314 ggcacacagcttggctgctgctgaggaagagccactgctatcaatgtaatg 373
 472 aactgtaagatgaggaagcttcttcaagagagccatgaagaagagagagag 531
 374 gactgcaagagctgcaagagcttcttccgagagctgctgctgctgctgctg 433
 532 tgcctcctcgag 588
 434 tatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 493
 589 gctcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 648
 494 cagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 553
 649 gaggcctgag 708
 554 gaac 613
 709 caac 768
 614 cactgac 666
 769 gac 828
 667 ttcctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 682
 829 gaggcctgag 888
 683 gaggcctgag 742
 889 gctcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 946
 743 ctatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 802
 947 tgcgag 1006
 803 cagcctgag 862

1007 agatctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 106
 863 caa-----cgcttggccacacttcacagagagagagagagagagagagag 914
 1067 tcaacttgcac 1126
 915 tggacttgccttgcac 974
 1127 tcaacttgcac 1186
 975 tccctgag 1044
 1187 agac 1240
 1035 ac 1094
 1241 gctcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1300
 1095 cagcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1154
 1301 tgcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1360
 1155 tgcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1214
 1361 gctcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1420
 1215 gctcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1274
 1421 tgaacttgcac 1480
 1275 tgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1334
 1481 tcaacttgcac 1540
 1335 tgaacttgcac 1398
 1541 gctcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1600
 1389 ccttgcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1448
 1601 gctcgcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1630
 1449 actgag 1478

RESULT 9
 US-08-649-619H-1
 Sequence 1, Application US/08649619H
 Patent No. 5871916
 GENERAL INFORMATION:
 APPLICANT: NAKAMURA, YOSUKE
 APPLICANT: SAITO, HIROKO
 TITLE OF INVENTION: ECDN PROTEIN AND DNA
 TITLE OF INVENTION: ENCODING THE SAME
 NUMBER OF SEQUENCES: 11
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLYNN, THIEL, ROUTELL &
 ADDRESSEE: TANIS, P.C.
 STREET: 2026 Rambling Road
 CITY: Kalamazoo
 STATE: Michigan
 COUNTRY: USA
 ZIP: 49008-1699
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inches, 1.44 MB
 MEDIUM TYPE: Storage
 COMPUTER: IBM PC/XT/AT compatible
 OPERATING SYSTEM: MS-DOS 5.0
 SOFTWARE: WordPerfect 5.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/649,619H
 FILING DATE:
 CLASSIFICATION: 435

RESUME 10
US OR 340-518 1
SEQUENCE 1, Applicant from US/08440518
? Patient No. 5607467
? GENERAL INFORMATION:
? APPLICANT: Friedman, Eitan
? APPLICANT: Hoffmann, M. Ralf
? APPLICANT: Rodan, Gidon
? APPLICANT: Schulman, David

[illegible]

RESUL 11
115-08-430-284-


```

QY 338 ctgaacttgtaacatgtgaaacaaatcttcttcttgaagaacccatgttcaacgag 397
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 CAGACTGGGTTCATCCCATGCTCCGAAAGCAACGACAGGCGCAAGCAAGGCGCCAG 476
QY 398 atgaagaatcggaatgcccaatctgcaatgtatgtggagacaagcgccactgtatc 457
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 CCGCAAGATGCTGGCCAGCAAGCTTGGCGTCTCTGTGGGCAACAGGCTCCGGCTTGC 536
QY 458 actcaatgtcatatgacatgtgaagaaatgaagagctttttcagagagccatgaacga 517
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 ACTACAACTGCTCAGCTGCGAAGGCTGCAAGGGCTTCTCCGGCGCAGTGTGTCGG 596
QY 518 aqcccgctgagtgagtccttcccgaaagcg---cccgagatcacccgaaagacc 574
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 GTGGGGCCAGGCGCTATGCTGTGCGGGGTGCCGAACCTGCCAAGATGAGAGCTTTTATGC 656
QY 575 ggcgaacagtgcgaagcctgcgcctgcaatgtcctgagagcgagcatgaagaagaga 634
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 657 GGGCAAGTGGCAGCAGCTGCGGGCTGCCAGATGCAAGGAGGAGGATGAGGAGCAGT 716
QY 635 tgaatgtcgaagagcgctgtgagagagcgagccttgatcaagcgaaagaaatg 694
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 717 GCTGCTTTCTGAGAGCAAGATCCGGAAGAGAGATTCGGAACGACGACGACGAGT 776
QY 695 aagcagagagactgcacactgagagtgcaq 726
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 777 CACAGTCACAGTGCAGTCACTGTGGGCGC 808

```

RESULT 15

PCT-US94-12883-3

Sequence 3, Application PC/TUS9412883

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR: COMPOSITIONS AND

TITLE OF INVENTION: METHODS

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: United States of America

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII

SOFTWARE: PatentIn Release #1.0, Version

SOFTWARE: #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/12883

FILING DATE: Concurrently Herewith

CLASSIFICATION:

APPLICATION NUMBER: US 08/152,003

FILING DATE: 10-NOV-1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: BARBARA S. KITCHELL

REGISTRATION NUMBER: 33,928

REFERENCE/POCKET NUMBER: ARCD154P--

TELEPHONE: (512) 418-3000

TELEPHONE: (713) 789-2679

TELEPHONE: 79-0924

INFORMATION FOR SEQ. ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1813 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
PCT-US94-12883-3

Query Match 4.5%; Score 97.2; DB 5; Length, 1813;
Best Local Similarity 58.8%; Pred. No. 2,76-14;
Matches 187; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

```

QY 412 gatcccaaatctgcgtatgtgagagaaagccactgtatcatatgaatg 471
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 229 GGCACAGAGCTTTGCGGTCTGTGTGGGCAAGGCGCTCCGCTTCCACTACAACTGTGCT 288
QY 472 acatgtgaagagtgcaagagcttttcagagagcgcatgaacgcaagcgctgag 531
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 289 AGCTGGAAGGCTGCAAGGCTTCTTCCGGCGCAGTGTGCTGTGCTGTGGAAGGCGC 348
QY 532 tgccttccggaagagcg---cctgagagatcacccgaaagcccgagcaatgcaq 588
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 TATGCTTGGCGGGTGGCGGAACCTGCCAGATGAGAGCTTTTCATGCGGCGCAAGTGGCAG 408
QY 589 gccctgcctgcagcaatgtcctgagagcgagcatgaagaaagaaatgaatgagac 648
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 409 CAGTCCAGCTGCGCAAGTGCAGAGGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAG 468
QY 649 gagcagagagagagagagcgcttgatcaagcgaaagaaatgaacgcaagagact 708
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 GAACAGATCCGAGAGAGAGATTCGGAACAGCAGCAGCAGCAGCAGCAGCAGCAGTGC 528
QY 709 caagcagagagagagagag 726
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 529 CAGTCACTGTGGGCGC 546

```

Search completed: July 11, 2001, 16:30:20
Job time: 14255 sec

J

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 11, 2001, 09:10:30 ; Search time 13.14 seconds

(without alignments)
1079.282 Million cell updates/sec

Title: US-09-276-935B-14

Perfect score: 2187
1 LEVPRKESWNRADPHCEDT.....AMLIPTATPLMDELFCITGS 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2165	99.0	434	1	PXR_HUMAN
2	1675.5	76.6	431	1	PXR_MOUSE
3	1656.5	75.7	431	1	PXR_RAT
4	774	35.4	422	1	VDR_XENLA
5	768.5	35.1	448	1	VDR_CORJA
6	768	35.1	451	1	VDR_CHICK
7	762	34.8	423	1	VDR_RAT
8	753.5	34.5	422	1	VDR_MOUSE
9	750	34.3	424	1	VDR_HVIN
10	750	34.3	427	1	VDR_HUMAN
11	731	33.4	348	1	NR13_HUMAN
12	685.5	31.3	358	1	NR13_RAT
13	680	31.1	358	1	NR13_MOUSE
14	472	21.6	461	1	NRH2_HUMAN
15	469.5	21.5	445	1	NRH2_MOUSE
16	463	21.2	446	1	NRH2_MOUSE
17	461.5	21.1	445	1	NRH3_RAT
18	461	21.1	446	1	NRH3_RAT
19	456.5	20.9	447	1	NRH3_HUMAN
20	431.5	19.7	395	1	THB_PAROL
21	428.5	19.6	757	1	ECR_LUCCU
22	426	19.5	369	1	THR_CHICK
23	426	19.5	461	1	THR1_HUMAN
24	425	19.4	373	1	THB_RANCA
25	424	19.4	373	1	THB_XENLA
26	423.5	19.4	675	1	ECR_AEDAE
27	423	19.3	476	1	THB2_HUMAN
28	419	19.2	414	1	THB2_XENLA
29	417	19.1	411	1	THB1_SHEEP
30	417	19.1	461	1	THB1_MOUSE
31	417	19.1	475	1	ECR_MOUSE
32	413.5	18.9	878	1	ECR_DROME
33	413	18.9	461	1	THB1_RAT

ALIGNMENTS

RESULT	ID	STANDARD	PKT	434 AA.
1	PXR_HUMAN			
AC	075469			
DT	15-JUL-1999 (Ref. 38, Created)			
DT	15-JUL-1999 (Ref. 48, Last sequence update)			
DT	01-OCT-2000 (Ref. 40, Last annotation update)			
DE	ORPHAN NUCLEAR RECEPTOR PXR (PREGNANE X RECEPTOR)			
DE	RECEPTOR PXR1).			
CN	NR112 OR PXR			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_Taxid:9606;			
HN	[1]			
RE	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE-98395173; PubMed-9727070;			
RA	Lehman J.M., McKee D.D., Watson M.A., Willson T.M., Moore J.T.,			
RA	Kliwer S.A.;			
RT	"The human orphan nuclear receptor PXR is activated by compounds that			
RT	regulate CYP3A4 gene expression and cause drug interactions.";			
RL	J. Clin. Invest. 102:1016-1023(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE-98443350; PubMed-9770465;			
RA	Bertilsson G., Heidrich J., Svensson K., Asman M., Jendryak L.,			
RA	Sydow-Backman M., Ohlsson R., Postlund H., Blomquist P.;			
RA	Berkensam A.;			
RT	"Identification of a human nuclear receptor defines a new signaling			
RT	pathway for CYP3A induction.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:12208-12213(1998).			
CC	- FUNCTION: ORPHAN RECEPTOR; ITS NATURAL LIGAND IS PROBABLY			
CC	PREGNANE. BINDS TO A RESPONSE ELEMENT IN THE CYP3A4 GENE PROMOTER.			
CC	- SUBUNIT: FORMS A HETERODIMER WITH RXR.			
CC	- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
CC	- TISSUE SPECIFICITY: EXPRESSED IN LIVER, COLON, AND SMALL			
CC	INTESTINE.			
CC	- INDUCTION: ACTIVATED BY NATURALLY OCCURRING STEROIDS SUCH AS			
CC	PREGNANOLONE AND PROGESTERONE.			
CC	- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.			
CC	NR1 SUBFAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC				
DR	EMBL: AF061056; AAD05446.1;			
DR	EMBL: AF084645; AAC64558.1;			
DR	MTM: 603065;			
DR	InterPro: IPR000424;			

QY 303 RLSYCLDITAGGFOQLLEPMUKFHYMLKQLDHEEYVLMQALSLSPDRPGVLOHRYV 362
 DB 300 RLSYCLDITAGGFOQLLEPMUKFHYMLKQLDHEEYVLMQALSLSPDRPGVLOHRYV 359
 QY 363 DOLODFPAITLKSYLECNPQPAHRLFLKIMAMITE-----FAT 402
 DB 360 DOLODFPAITLKSYLECNPQPAHRLFLKIMAMITE-----FAT 419
 QY 403 PLMOELFGIT 412
 DB 420 PLMOELFST 429
 RESULT 3
 PXR_RAT STANDARD: PRT: 431 AA.
 ID PXR_RAT
 AC 09R1A7:
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ORPHAN NUCLEAR RECEPTOR PXR (PREGNANE X RECEPTOR).
 GN NR1I2 OR PXR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RX MEDLINE=99345883; PubMed=10415106;
 RA Zhang H., Lecuysse E., Liu L., Hu M., Matoney L., Zhu W., Yan B.;
 RT "Rat pregnane X receptor: molecular cloning, tissue distribution, and
 xenobiotic regulation." 368:14-22(1999).
 RL Arch. Biochem. Biophys. 368:14-22(1999).
 CC -1- FUNCTION: ORPHAN RECEPTOR; ITS NATURAL LIGAND IS PROBABLY
 PREGNANE. BINDS TO A RESPONSE ELEMENT IN CYP3A GENES PROMOTER (BY
 SIMILARITY).
 CC -1- SUBUNIT: FORMS A HETERODIMER WITH RXR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NRI SUBFAMILY.
 CC -----
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib.ch).
 CC -----
 DR EMBL: AF151377; AA047214.1;
 DR InterPro: IPR000324;
 DR InterPro: IPR000536;
 DR InterPro: IPR001628;
 DR InterPro: IPR001723;
 DR Pfam: PF00104; hormone_rec.1.
 DR Pfam: PF00105; zf-C4.1.
 DR PRINTS: PR00047; STEROIDFINGER.
 DR PRINTS: PR00350; VITAMIND.
 DR PRINTS: PR00398; STRDHOMONER.
 DR PROSITE: PS00031; NUCLEARRECEPTOR.1.
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KM Zinc-finger.
 FT DNA_BIND 38 104 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 38 58 C4-TYPE.
 FT ZN_FING 74 99 C4-TYPE.
 FT DOMAIN 105 201 HINGE.
 FT DOMAIN 202 431 LIGAND-BINDING.
 FT SEQUENCE 431 AA; 49660 MW; 4B545F21F9439697 CRC64;

Query Match 75.7%; Score 1656.5; DR 1: Length 431;
 Best Local Similarity 72.8%; Pred. No. 4, 9e-126;

Matches 313; Conservative 41; Mismatches 55; Indels 21; Gaps 2;
 QY 3 VPKRSNNAHADVYNGEDTHESVPKRSVNADEVGSPQICVGRKATGYHNNWTCECK 62
 DB 1 MKPEERMMHVGILVOREADSVLEEP-INVDEEDGSLQICVGRKANCVHNNWTCECK 59
 QY 63 GFFRAMRNARLRCPPFKGACETTRKTRPOCAGRLKCLIESIMKRMKMSDAVEER 122
 DB 60 GFFRAMRNARLRCPPFKGACETTRKTRPOCAGRLKCLIESIMKRMKMSDAVEER 119
 QY 123 ALIKRKSERTQPLVGVGLTEVRMMIRELMDQMPTPTSHFNRLPYVLSNG 182
 DB 120 ALIKRKSERTQPLVGVGLTEVRMMIRELMDQMPTPTSHFNRLPYVLSNG 179
 QY 183 ELPELIQAPSRPAKMSQVKKDLSLKVSLQREDSVNNYKPADSCKEIFSLPH 242
 DB 180 ELPELIQAPSRPAKMSQVKKDLSLKVSLQREDSVNNYKPADSCKEIFSLPH 239
 QY 243 MADNSTYFKGISFAKISYERDLPIEDQISLKGAFELQRLNFTVNAEIGTWNG 302
 DB 240 MADNSTYFKGISFAKISYERDLPIEDQISLKGAFELQRLNFTVNAEIGTWNG 299
 QY 303 RLSYCLDITAGGFOQLLEPMUKFHYMLKQLDHEEYVLMQALSLSPDRPGVLOHRYV 362
 DB 300 RLSYCLDITAGGFOQLLEPMUKFHYMLKQLDHEEYVLMQALSLSPDRPGVLOHRYV 359
 QY 363 DOLODFPAITLKSYLECNPQPAHRLFLKIMAMITE-----FAT 402
 DB 360 DOLODFPAITLKSYLECNPQPAHRLFLKIMAMITE-----FAT 419
 QY 403 PLMOELFGIT 412
 DB 420 PLMOELFST 429
 RESULT 4
 VDR_XENLA STANDARD: PRT: 422 AA.
 ID VDR_XENLA
 AC 013124;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE VITAMIN D3 RECEPTOR (VDR) (1,25-DIHYDROXYVITAMIN D3 RECEPTOR).
 GN VDR OR NR1I1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=97307679; PubMed=9165021;
 RA Li Y.C., Bergwitz G., Joepner H., Demay M.B.;
 RT "Cloning and characterization of the vitamin D receptor from Xenopus
 laevis." 138:2347-2353(1997).
 RL Endocrinology 138:2347-2353(1997).
 CC -1- FUNCTION: VDR MEDIATES THE ACTION OF VITAMIN D3 BY CONTROLLING
 THE EXPRESSION OF HORMONE SENSITIVE GENES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: DETECTED IN ALL THE TISSUE EXAMINED. HIGHEST
 LEVEL IN SMALL INTESTINE AND SKIN.
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT STAGE 13. INCREASES
 GRADUALLY AND PEAKS AT STAGE 57-61 THEN DECREASES TO THE LEVEL
 SEEN IN ADULT.
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NRI SUBFAMILY.
 CC -----
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its


```

D6 307 WTGSGNDKXKYSVDTQAGHSMLEPLVKQVGLKTLNHEEHVLLMATICILSPBPG 366
QY 356 VIGHVVOLOLOPAILTKSYTECKRPOPAHFLLKIMAMLET----- 400
D6 367 VQDTSLVSIQDRSDTLQTYLRCHRPSPGLYAKMKIQLADLRSLNEHSKQYRCIS 426
QY 401 -----ATPLMOELFG 410
D6 427 FQPEHSMOLTPLVLEFG 444

RESULT 6
VDR_CHICK
ID VDR_CHICK STANDARD: PRT: 451 AA.
AC Q4292:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VITAMIN D3 RECEPTOR (VDR) (1,25-DIHYDROXYVITAMIN D3 RECEPTOR).
GN VDR OR NR111
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEGHORN; TISSUE-Kidney;
RX MEDLINE=97223369; PubMed=9056239;
RA Lu Z., Hanson K., Deluca H.F.;
RT "Cloning and origin of the two forms of chicken vitamin D receptor.";
RL Arch. Biochem. Biophys. 339:99-106(1997).
RN [2]
RP SEQUENCE OF 45-114 FROM N.A.
RX MEDLINE=87149040; PubMed=3029866;
RA McDonnell D.P., Mangelsdorf D.J., Pike J.W., Haussler M.R.,
RA O'Malley B.W.;
RT "Molecular cloning of complementary DNA encoding the avian receptor
RT for vitamin D.";
RL Science 235:1214-1217(1987).
CC -1- FUNCTION: VDR MEDIATES THE ACTION OF VITAMIN D3 BY CONTROLLING
CC THE EXPRESSION OF HORMONE SENSITIVE GENES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) ARE PRODUCED BY USE OF
CC ALTERNATIVE INITIATION CODONS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY AND INTESTINE.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NRI SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
D6 EMBL: AF011356; AA62579.1; -
D6 InterPro: IPR000324; -
D6 InterPro: IPR000536; -
D6 InterPro: IPR001628; -
D6 Pfam: PF00104; hormone_rec; 1
D6 Pfam: PF00105; zf-c4; 1
D6 PRINTS: PR00047; STEROIDFINGER.
D6 PRINTS: PR00350; VITAMINDR.
D6 PROSITE: PS00031; NUCLEAR_RECEPTOR; 1
D6 Receptor: Transcription regulation; DNA-binding; Nuclear protein;
D6 Zinc-finger.
D6 CHAIN 1 451 VITAMIN D3 RECEPTOR, FORM A.
D6 FT 15 451 VITAMIN D3 RECEPTOR, FORM B.
D6 INIT_MET 15 15 FOR FORM B.

```

```

FT DNA_BIND 47 112 NUCLEAR RECEPTOR TYPE.
FT ZN_FING 47 67 C4-TYPE.
FT ZN_FING 83 107 C4-TYPE.
FT DOMAIN 113 215 HINGE.
FT DOMAIN 216 451 LIGAND-BINDING.
SU SEQUENCE 451 AA: 51299 MW: 20786B6ACBDBE5FC CCK44;

Query Match 35.1% Score 768; Dh 1; Length 451;
Best Local Similarity 38.0%; Pred. No. 166-54;
Matches 172; Conservative 71; Mismatches 148; Indels 62; Gaps 10;

QY 8 SWNH-----ADVEHCEDTSVRKPSVNADEYVGQICRGVGNATGTHFVMT 57
D6 7 SMDPQOQSMAYIPADMDTYAASTSLP-DPAGDPDRNV-PRICVGVDPATGFHFAMT 63
QY CEGCKGFFRRAMKRNARLCPFRKACETLTKTRQOCACRKLCSKKKREKMSIFA 117
D6 64 CEGCKGFFRRAMKRNARLCPFRKACETLTKTRQOCACRKLCSKKKREKMSIFA 122
QY 118 VEERRALIKRKRSERTGTPLRGVGLTEPQKMLREMDAMKTPDTTSHKKNKLP- 175
D6 123 VQRRKEMILKKREDEALKEKSLPK-LSEQCKQVIDITLLEAHIKTPDTTYSDFKFRPVR 181
QY 176 -----GVLSGCHLPESIOAPSPREAKMSQVKRDKICLKVSIQIKRDKSW 223
D6 182 SKFSRHMAIHSSSVASQDSESDSDYGVSDAFAPFPMYDMSNLDSEESDESPSM 241
QY 224 NYKPPADSGKEIFSLIPHMADSMYMPKGIISAKVSYVRDLPPIHOULSLKGAPEL 283
D6 242 NIELPH-----LPMPLHADLVSYQKYIGAKMKPGKROLIAHQULAKSSALIV 294
QY 284 COLRPNTVNAETGTWEGCR--ISYGLD--TAGSPQOULLFPMILKHYMKLKLHBEV 340
D6 295 IMLSNOSTWMEQMSWTCGSDNFYKVSQVDTQAGHSMLEPLVKQVGLKTLNHEEH 354
QY 341 VLMAISLSPSPRGVQGHRRVVDLOEOGFAITKSYTECKRPOPAHFLLKIMAMLET 400
D6 355 VLMAICILSPDRGVQDTSVESTIQLRSDILOTLYLRCHRPSPGLYAKMKIQLADL 414
QY 401 -----ATPLMOELFG 410
D6 415 RSLNEHSKQYRCISLSPQPHSMOLTPLVLEFG 447

RESULT 7
VDR_RAT
ID VDR_RAT STANDARD: PRT: 423 AA.
AC P13053:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VITAMIN D3 RECEPTOR (VDR) (1,25-DIHYDROXYVITAMIN D3 RECEPTOR).
GN VDR OR NR111.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89071726; PubMed=2849110;
RA Burmester J.K., Wiese R.J., Maeda N., Deluca H.;
RT "Structure and regulation of the rat 1,25-dihydroxyvitamin D3
RT receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9499-9502(1988).
RN [2]
RP SEQUENCE OF 58-423 FROM N.A.
RX MEDLINE=88124963; PubMed=2829212;
RA Burmester J.K., Maeda N., Deluca H.F.;
RT "Isolation and expression of rat 1,25-dihydroxyvitamin D3 receptor
RT cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:1005-1009(1988).
CC -1- FUNCTION: VDR MEDIATES THE ACTION OF VITAMIN D3 BY CONTROLLING;

```


DB 199 1M -DL-----NEEGS-----DDPSVTLDELSPMLHLADLVSYGKVIQFAPKM 242

QY 261 ISYFQDLPIDQISLLKGAAPFLCQLRNTVFNACTGMEGRSLSYCLEPDT---AGGFQO 317

DB 243 IGFHFDLSDDDQIVLLKSSALFVIMLRNSQSFDTMDMSGSGQYKDYIDTVSRAGHTL 302

QY 318 LLEMLFHWMLKLOLHEEFYVIMQAI SLSPDRPGVLOHRVVDQLOEQFAITLKSYSI 377

DB 303 ELIEPLIFQVGLKKNLHEEHEVILMAICIVSPDRPGVDQAKLVEALIDRLSNLTQYI 362

QY 378 ECRNQPAPHRFLFKIMAMLTFF-----ATPLMOELFG 410

DB 363 RCRHRRPSSHQYAKMIQKIALDRSLNEHNSQYSLSHFQENSMKLTPLVEVVG 418

RESULT 9

ID VDR_BOVIN STANDARD: PRT: 424 AA.

AC Q28037:

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE VITAMIN D3 RECEPTOR (VDR) (1,25-DIHYDROXYVITAMIN D3 RECEPTOR).

GN VDR OR NR111.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP MEDLINE=97034797; PubMed=8880453;

RA Nelbers H.L., Bosworth B.T., Reinhardt T.A.;

RT "Nucleotide sequence of the bovine vitamin D3 receptor.";

RL J. Dairy Sci. 79:1313-1315(1996).

CC - FUNCTION: VDR MEDIATES THE ACTION OF VITAMIN D3 BY CONTROLLING

CC THE EXPRESSION OF HORMONE SENSITIVE GENES.

CC - SUBCELLULAR LOCATION: NUCLEAR

CC - DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,

CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.

CC - SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.

CC NRI SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb.ch/announce/>

CC or send an email to license@isb.ch).

CC -----

DR EMBL: U50200; AAH01543.1; .

DR HSSP: P03372; IMCQ.

DR InterPro: IPR000324; .

DR InterPro: IPR00036; .

DR InterPro: IPR001628; .

DR Pfam: PF00104; hormone-rec; 1.

DR Pfam: PF00105; zf-C4; 1.

DR PRINTS: PR00047; STEROIDRECEPTOR.

DR PRINTS: PR00350; VITAMINR.

DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.

DR Receptor: Transcription regulation; DNA-binding; Nuclear protein;

KM Zinc-finger; Phosphorylation.

FT DNA_BIND 21 86 NUCLEAR RECEPTOR-TYPE.

FT ZN_FING 21 41 C4-TYPE.

FT ZN_FING 57 81 C4-TYPE.

FT DOMAIN 87 188 HINCE.

FT DOMAIN 189 424 LIGAND-BINDING.

SO SEQUENCE 424 AA; 47957 MW; E3E24926CE38CB7D CRC64;

Query Match 34.3%; SCOP 750; DB 1; Length 424;

Best Local Similarity 40.3%; Pfam No. 4.2e-53;

Matches 166; Conservative 62; Mismatches 136; Indels 48; Gaps 8;

QY 38 POLRVCQAKAGVHFNVMTCGGKGFRRAMKNNALRCKPFRKACDIIJKTROQAO 97

DB 18 PRIGVCCDRAFGHPNMTCEGCKGFFRRSMKRALFTGPF-NGDRITIKDNKHQAC 76

QY 98 RLKCLSGMKKEMIMSPAVEERBALIKRKSKRTGTQPLGVGLTEQBMHREIMDA 157

DB 77 RLKRCVIGIMKKKFTLTDEEVGRREMLIKREERALKDLSLRK-LSEFQRIIATLIDA 135

QY 158 QMKTFDTFFSFFKFNRLPGVLSG-----CELPESIQASREAKKWSQVKDLA----- 207

DB 136 HAKTYDPTSPFCQPRPVARNDDGGSHPSKPNKRHTPS--FSDSSSSGSDRHTSSGM 193

QY 208 ---SLKVSIGLRKDGGSVWNYKPPADSGKEIFSLPHMAIMSTYMRGIIIFAKVISTF 264

DB 194 MSSSFNMLDSEEDSD-----DPSVTLLESLQISMLPHLADLVSYGKVIQFAPKM 248

QY 265 KDLPLIEDQISLLKGAAPFLCQLRNTVFNACTGMEGRSLSYCLEPDTAGGVLLE 321

DB 249 KDLTSEDDQIVLLKSSALFVIMLRNSQSFDTMDMSGSGQYKDYIDTVSRAGHTL 308

QY 322 PMLKFWMLKLOLHEEYVIMQAI SLSPDRPGVLOHRVVDQLOEQFAITLKSYSI 381

DB 309 PLIKFOVALKKNLHEEHEVILMAICIVSPDRPGVDQAKLVEALIDRLSNLTQYI 368

QY 382 PUPAHRFLFKIMAMLTFF-----ATPLMOELFG 410

DB 369 PPGSHLLYAKMIQKIALDRSLNEHNSQYSLSHFQENSMKLTPLVEVVG 420

RESULT 10

ID VDR_HUMAN STANDARD: PRT: 427 AA.

AC P11473;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE VITAMIN D3 RECEPTOR (VDR) (1,25-DIHYDROXYVITAMIN D3 RECEPTOR).

GN VDR OR NR111.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP MEDLINE=88217887; PubMed=2835767;

RA Baker A.R., McDonnell D.P., Hughes M., Crisp T.M., Mangelsdorf D.J.,

RT Hausler M.R., Pike J.W., Shine J., O'Malley B.W.;

RT "Cloning and expression of full-length cDNA encoding human vitamin D

RT receptor.";

RT Proc. Natl. Acad. Sci. U.S.A. 85:3294-3298(1988).

RN [2]

RP MEDLINE=92374083; PubMed=1324736;

RA Goto H., Chen K.S., Prahl J.M., Deluca H.F.;

RT "A single receptor identical with that from intestine/147D cells

RT mediates the action of 1,25-dihydroxyvitamin D-3 in HL-60 cells.";

RL Biochim. Biophys. Acta 1132:103-108(1992).

RN [3]

RP MEDLINE=97355582; PubMed=9212063;

RA Miyamoto K., Kesterson R.A., Yamamoto H., Takeda E., Nishiwaki F.,

RT Takeda S., Inoue Y., Morita K., Takeda E., Pike J.W.;

RT "Structural organization of the human vitamin D receptor chromosomal

RT gene and its promoter.";

RL Mol. Endocrinol. 11:1165-1179(1997).

RN [5]

SEQUENCE OF 24 90 FROM N.A.
 TISSUE: Peripheral blood;
 MEDLINE 91210272; PubMed 1870412;
 RA Yu X. F., Morhalla H., Hoshinger F.G., Mulonias S.C.;
 "Vitamin D receptor expression in human lymphocytes. Signal
 transduction and characterization by western blots and DNA
 sequencing";
 RA J. Biol. Chem. 266:7588-7595(1991).
 RA [6]
 RA VARIANTS ASP 43 AND GLN 73;
 RA MEDLINE 90072761; PubMed 2849209;
 RA Hughes M.R., Malloy P.J., Kiback D.G., Kesterson K.A., Pike J.W.,
 Feldman D., O'Malley B.W.;
 "Point mutations in the human vitamin D receptor gene associated with
 hypocalcemic rickets";
 RA Science 242:1702-1705(1988).
 RA [7]
 RA VARIANT GLN 15;
 RA MEDLINE 9116243; PubMed 8481803;
 RA Yagi H., Ozono K., Miyake H., Nagashima K., Kurokawa T., Pike J.W.;
 "A new point mutation in the deoxyribonucleic acid binding domain of
 the vitamin D receptor in a kindred with hereditary
 1,25-dihydroxyvitamin D-resistant rickets";
 RA J. Clin. Endocrinol. Metab. 76:509-512(1993).
 RA [8]
 RA VARIANT GLN 50;
 RA MEDLINE 9135398; PubMed 1652893;
 RA Saito T., Ito M., Takeda F., Mahnubul Hoo A.H.M., Naito E., Yokota T.,
 Sato T., Pike J.W., Kuroda Y.;
 "A unique mutation in the vitamin D receptor gene in three Japanese
 patients with vitamin D-dependent rickets type II: utility of single-
 strand conformation polymorphism analysis for heterozygous carrier
 detection";
 RA Am. J. Hum. Genet. 49:668-674(1991).
 RA [9]
 RA VARIANT GLN 80;
 RA MEDLINE 9112570; PubMed 2177843;
 RA Sore T., Marx S.J., Liberman M.A., Pike J.W.;
 "A unique point mutation in the human vitamin D receptor chromosomal
 gene confers hereditary resistance to 1,25-dihydroxyvitamin D3";
 RA Mol. Endocrinol. 4:623-631(1990).
 RA [10]
 RA VARIANT GLN 80;
 RA MEDLINE 94145129; PubMed 8106618;
 RA Malloy P.J., Weisman Y., Feldman D.;
 "Hereditary 1,25-dihydroxyvitamin D-resistant rickets resulting
 from a mutation in the vitamin D receptor deoxyribonucleic
 acid binding domain";
 RA J. Clin. Endocrinol. Metab. 78:313-316(1994).
 RA [11]
 RA VARIANT LEU 274;
 RA MEDLINE 93316633; PubMed 8792085;
 RA Kristjansson K., Kot A.R., Hewison M., O'Riordan J.L.H., Hughes M.R.;
 "Two mutations in the hormone binding domain of the vitamin D
 receptor cause tissue resistance to 1,25-dihydroxyvitamin D3";
 RA J. Clin. Invest. 92:12-16(1994).
 RA [12]
 RA VARIANTS GLN 45 AND ILE 47;
 RA MEDLINE 95129267; PubMed 7828446;
 RA Kot A.R., Hewison M., Kristjansson K., Inai H., Hughes M.R.,
 O'Riordan J.L.H.;
 "Two mutations causing vitamin D resistant rickets: modeling on the
 basis of steroid hormone receptor DNA-binding domain crystal
 structures";
 RA J. Biol. Chem. 269:41581-41590(1994).
 RA [13]
 RA VARIANT ASP 46;
 RA MEDLINE 96272879; PubMed 8675579;
 RA Liu H. T., Malloy P.J., Sakai N., Al-Ashkar A., Feldman D.;
 "A novel mutation in the deoxyribonucleic acid-binding domain of the
 vitamin D receptor causes hereditary 1,25-dihydroxyvitamin D resistant
 rickets";
 RA J. Clin. Endocrinol. Metab. 81:2564-2569(1996).

[14]
 RA VARIANTS SER 114 AND CYS 191;
 RA MEDLINE 97120600; PubMed 8961271;
 RA Gallilei G.K., Seizova S.H., Hausloy C.A., Hsieh J. C.,
 Callahan M.A., Juritka P.W., Thompson P.D., Lee S.M., Zetser J.E.,
 Haussler M.R.;
 "Vitamin D receptors from patients with resistance to 1,25-
 dihydroxyvitamin D(3): point mutations confer reduced transactivation
 in response to ligand and impaired interaction with the retinoid X
 receptor heterodimeric partner";
 RA Mol. Endocrinol. 10:1617-1631(1996).
 RA [15]
 RA VARIANT GLN 105;
 RA MEDLINE 97158648; PubMed 9007998;
 RA Malloy P.J., Evertsall T.R., Gross C., van Malderom L., Paulson R.,
 Feldman D.;
 "Hereditary vitamin D resistant rickets caused by a novel mutation in
 the vitamin D receptor that results in decreased affinity for hormone
 and cellular hyporesponsiveness";
 RA J. Clin. Invest. 99:297-304(1997).
 RA [16]
 RA FUNCTION: VDR MEDIATES THE ACTION OF VITAMIN D3 BY CONTROLLING
 THE EXPRESSION OF HORMONE SENSITIVE GENES.
 RA [17]
 RA SUBCELLULAR LOCATION: NUCLEAR.
 RA [18]
 RA DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
 A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
 RA [19]
 RA DISEASE: DEFECTS IN VDR ARE THE CAUSE OF TYPE IIA RICKETS (ALSO
 KNOWN AS HYPOCALCAEMIC VITAMIN D-RESISTANT RICKETS (HVRDR)). HVRDR
 IS MOST FREQUENTLY AN AUTOSOMAL RECESSIVE DISORDER CHARACTERIZED
 BY SEVERE RICKETS, HYPOCALCAEMIA AND SECONDARY HYPERPARATHYROIDISM.
 RA [20]
 RA SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 RA [21]
 RA SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 RA [22]
 RA CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-4 IS THE INITIATOR.
 RA [23]
 RA THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 RA [24]
 RA EMBL: J03258; AAA61274.1;
 RA EMBL: X67482; CAA47824.1;
 RA EMBL: AF026260; AAB95155.1;
 RA EMBL: AF002168; BAAB3489.1;
 RA EMBL: AB002161; BAAB3489.1; JOINED.
 RA EMBL: AB002162; BAAB3489.1; JOINED.
 RA EMBL: AB002163; BAAB3489.1; JOINED.
 RA EMBL: AB002164; BAAB3489.1; JOINED.
 RA EMBL: AB002165; BAAB3489.1; JOINED.
 RA EMBL: AB002166; BAAB3489.1; JOINED.
 RA EMBL: AB002167; BAAB3489.1; JOINED.
 RA EMBL: M55208; AAA61274.1;
 RA PIR: S28200; A28200.
 RA PIR: S24174; S24174.
 RA HSSP: P03372; HRCV.
 RA TRANSFAC: T09885;
 RA MIM: 601769;
 RA MIM: 277440;
 RA InterPro: IPR000324;
 RA InterPro: IPR000536;
 RA InterPro: IPR001628;
 RA Pfam: PF00104; hormone_rec_1;
 RA Pfam: PF00105; 2f_c4_1;
 RA PRINTS: PR00047; STEROID-INGER.
 RA PRINTS: PR00350; VITAMINOR.
 RA PROSITE: PS00041; NUCLEAR-RECEPTOR_1;
 RA Kinase: Transcription regulation; DNA-binding; Nuclear protein;
 RA Zinc-finger; Phosphorylation; Disease mutation;
 RA DNA_BIND: 24 89
 RA ZN_FING: 24 44 C4 TYPE.
 RA ZN_FING: 60 84 C4 TYPE.
 RA DOMAIN: 90 191 HINGE.


```

FT DOMAIN 192 427 LIGAND-BINDING.
FT VARIANT 33 33 G->D (IN RICKETS).
FT VARIANT 35 35 /FTID-VAR.004656.
FT VARIANT 45 45 H->Q (IN RICKETS).
FT VARIANT 45 45 /FTID-VAR.004657.
FT VARIANT 45 45 K->E (IN RICKETS).
FT VARIANT 46 46 /FTID-VAR.004658.
FT VARIANT 46 46 G->D (IN RICKETS).
FT VARIANT 47 47 /FTID-VAR.004659.
FT VARIANT 47 47 F->I (IN RICKETS).
FT VARIANT 50 50 /FTID-VAR.004660.
FT VARIANT 50 50 R->Q (IN RICKETS).
FT VARIANT 73 73 /FTID-VAR.004661.
FT VARIANT 73 73 R->Q (IN RICKETS).

Query Match 34.3%, Score 750; DB 1; length 427;
Best Local Similarity 40.3%, Pred. No. 4.2e-53;
Matches 166; Conservative 62; Mismatches 136; Indels 48; Gaps 8;

QY 38 PQCVCVCGPKATGYHFNVTGCGKGFRRRAKRNALRCPRKGCATITKTRPGCQAC 97
DB 21 PRICVCGDRATGFFPNAITCGCGFFRRSKRKALFTTCTP-NCDCRITKDNRRHCQAC 79
QY 98 RLKLCLESQMKKEMIMSDAEVEERALLIKRKSEKRTGTQPLVCGLTDEORMIMRELMDA 157
DB 80 RLKRCVDIGMKKEPILTDEEVQKRREMLIKKKEEALDKSLRPK-LSEEQRIITALLDA 138
QY 158 QMKETDTTFSHKRNRLFGVLISSG----CELPESLQAPSREFAAKWSVKRDKC----- 207
DB 139 HKKIVDPYTSDFQCFPRPYRVNMGSGSHPSRPNRHTPS--FSCDSSSSCSHCITSSDM 196
QY 208 ---SLKVSITGJGEGSGSWNKKPPADSGKELFSLPHMADMSTYFMFGIISFAVISYF 264
DB 197 MDSSFSNLDLSEESD-----DPSVTLELSQSLMLPHLADIVSYSLCKVIGFAKMGPF 251
QY 265 RDLPLEDITSLKGAFFELCQAFRTVFNATGTWEGRLSY---CLPDIAGGFPQOLLLE 321
DB 252 RDLTSEDQVILKSSAIEVIMLRNSESFTMDMSMTGCGQDYKAVSVYTKAGHLELIE 311
QY 322 PMLKHYMKKQLQLEBEVEVLMQALISLSPDRPGVLOHNVQDLOEFAITLKSTIECNR 381
DB 312 PLKRYOVGKRLNLEBEVEVLMQALISLSPDRPGVLOHNVQDLOEFAITLKSTIECNR 371
QY 382 POPARFLFKIMALTEF-----APFLMDPLPG 410
DB 372 PPPGSHILAKMTQKLADRLSLNEHSGQYKCLSPQECSMKLTPLVLEVFG 423

RESULT 11
NR13_HUMAN STANDARD; PRT: 348 AA.
AC Q14994.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ORPHAN NUCLEAR RECEPTOR NR113 (CONSTITUTIVE ANDROSTANE RECEPTOR) (CAR)
DE (ORPHAN NUCLEAR RECEPTOR MB67).
GN NR113.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94158827; PubMed=8114692;
RA Baes M., Gulick T., Choi H.S., Martinoli M.G., Simba D., Moore D.D.;
RT "A new orphan member of the nuclear hormone receptor superfamily that
RT interacts with a subset of retinoic acid response elements."
RL Mol. Cell. Biol. 14:1544-1552(1994).
CC -!- FUNCTION: BINDS AND TRANSACTIVATES THE RETINOIC ACID RESPONSE
CC ELEMENTS THAT CONTROL EXPRESSION OF THE RETINOIC ACID RECEPTOR
CC BETA 2 AND ALCOHOL DEHYDROGENASE 3 GENES.

```

```

CC -!- SUBUNIT: HETERODIMER OF NR113 AND RXR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DB EMBL: Z30425; CAA83016.1;
DB MIM: 603881;
DB InterPro: IPR000324;
DB InterPro: IPR000536;
DB InterPro: IPR001628;
DB InterPro: IPR001728;
DB Pfam: PF00104; hormone_rec_1;
DB Pfam: PF00105; z1-C4; 1;
DB PRINTS: PR00047; STROIDPFINGER.
DB PRINTS: PR00350; VITAMINDER.
DB PRINTS: PR00546; THYROIDHORM.
DB PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Activator.
FT DNA_BIND 11 76 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 11 31 C4-TYPE.
FT ZN_FING 47 71 C4-TYPE.
SQ SEQUENCE 348 AA; 39558 MM; 92EAD07D7DCB9DA CIRC64;

```

```

Query Match 33.4%, Score 731; DB 1; length 348;
Best Local Similarity 40.5%, Pred. No. 1.1e-51;
Matches 157; Conservative 59; Mismatches 100; Indels 72; Gaps 6;

QY 41 CRVCGDKATGYHFNVTGCGKGFRRRAKRNALRCPRKGCATITKTRPGCQACRLR 100
DB 11 CVVCGDQATGYHFNALTCGCGKGFRRRTVSKSIGTCTPF-AGSCVSKTORRHCPACQLQ 69
QY 101 KLESQMKKEMIMSDAEVEERALLIKRKSEKRTGTQPLVCGLTDEORMIMRELMDQMK 160
DB 70 KCLDQMKKMDILSABEALBRARQAOORAOQTVQ-----LSKEDEFLITLLGHTK 123
QY 161 TFDITTFSHKRNRLFGVLISSGCELPESLQAPSREFAAKWSVKRDKLSIKVSLTGEFG 220
DB 124 HMGTMFEGVQFPPRAHLFTIHQ-PLPTLAP----- 153
QY 221 SWWNTKPPADSGKELFSLPHMADMSTYFMFGIISFAKVISYFROLDPEVQISLKGAA 280
DB 154 -----VLPLTVHFADINTFNVLQVTKFQKDLVPRSPLEQJSLKGA 198
QY 281 FELCOLRNTVFNATGTWEGRLSYCLEDETAG-GFQULLLEPMLKHYMKKQLQLEBE 339
DB 199 VEICHIVLNTFCQOTQNFLLCPRLTYTLEDGARVGFVEFLFHHGHTIRKQLQLEPE 258
QY 340 YLMQALISLSPDRPGVLOHNVQDLOEFAITLKSTYIECNRPAPARHFLFKIMALTE 399
DB 259 YVLLAAMALFSPDRPGVQYQREIDQLOEEMALTLQSYTIKGGQRPRPRFLYAKLLGLAE 318
QY 400 FAT-----PLMDPL 408
DB 319 LRSINEAVGYQIHIQGLSAMPIQEI 346

RESULT 12
NR13_RAT STANDARD; PRT: 358 AA.
AC U9081.
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)

```


FT DNA_BIND 21 86 NUCLEAR RECEPTOR-TYPE.
 FT ZN_BINDING 21 41 C4-TYPE.
 FT ZN_BINDING 57 81 C4-TYPE.
 FT VARSPLIC 281 286 BRPGET -> GRCMOS (IN ISOFORM CAR2).
 FT VARSPLIC 287 358 MISSING (IN ISOFORM CAR2).
 SO SEQUENCE 358 AA: 40894 MW: 4607730FE78CADBC CRC64:

Query Match 31.1%; Score 680; DB 1; Length 358;
 Best Local Similarity 37.7%; Fred. No. 1.5e-47;
 Matches 150; Conservative 58; Mismatches 118; Indels 72; Gaps 6;

31 ADEEFGPPIRCVCGDKATGTFHFNWTCGCKGFFRRAKRMARLRCPRKGCACETIKRT 90
 11 ADEEFGPPIRCVCGDKATGTFHFNWTCGCKGFFRRAKRMARLRCPRKGCACETIKRT 90
 91 PRQCGACLRKLCESGCKEMKMSDEAVERRALIRKKSEKGTQPLGCGI:TFRRMM 150
 70 RRHCPACRLQCKLNGMRKMDILSAEALALRRARQQRRAEKASQ-----LNGQCKEL 123
 151 IRFLMDAOKMTEDTTFSHKKNRIRGCVVSGCELPESIQARSREERAKMSVKRLDLSLK 210
 124 VQILGAHTRHVGPLEDFVDFKPPAYL----- 151
 211 VSLQRGEDGSVMYKPPADSGKEIFSLPHADMSYMFKGLISPAKVISYFRLDPE 270
 152 -----FMHHPPOPKG--PVLPLTLHPADINTFMVOJLTKFDLPFLSTIME 198
 271 DQISLKGAFELCOLRENTVNAETGMEGRLSYCEDTA-GGFOQLLEPMLEKPHYM 329
 199 DQISLKGAFELHLSINTFCIGTENFGCGPKYKMDAVHAPQVFFLESLIHFKN 258
 330 LKKLQHEEYVLMQALISFSPDPCVLOHRYVDLOEQFAITLKSYLECNRPQAHRL 389
 259 LKGLHQPPEYVIMATALLSPDRGVTOREEDLOEEMALINNHMEQSRLOSLREF 318
 390 FLKIMAMLETF-----ATPLMOEL 408
 319 YAKIMGLADLRISNNASYELQLELSAMTFLGEI 356

RESULT 14
 NRH2_HUMAN STANDARD: PKT: 461 AA.
 AC P55055: 012970; Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OXYSTEROLS RECEPTOR LXR-BETA (LIVER X RECEPTOR BETA) (NUCLEAR ORPHAN RECEPTOR LXR-BETA) (UBIQUITOUSLY-EXPRESSED NUCLEAR RECEPTOR) (NUCLEAR RECEPTOR LXR-BETA)
 GN NR1H2 OR LXR-B OR UNR OR NER.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Osteosarcoma;
 RA MEDLINE=95011628; PubMed=7926814;
 RA Shinar D.M., Endo N., Rutledge S.J., Vogel R., Rodan G.A., Schmidt A.;
 RT "NER, a new member of the gene family encoding the human steroid hormone nuclear receptor."
 RT Gene 147:273-276(1994).
 RN [2]
 RP SEQUENCE OF 7-461 FROM N.A.
 RA Song C., Konkantis J.M., Hlipakka R.A., Liao S.;
 RT "ubiquitous receptor: a novel receptor that modulates gene activation by retinoic acid and thyroid hormone receptors."
 RT Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 CC -i- FUNCTION: ORPHAN RECEPTOR. BINDS PREFERENTIALLY TO DOUBLE-STRANDED OLIGONUCLEOTIDE DIRECT REPEATS HAVING THE CONSENSUS HALF-SITE SEQUENCE: 5'-AGGTCA-3' AND 4-NT SPACING (DR-4).

CC -i- SUBUNIT: FORMS A HETERODIMER WITH RXR.
 CC -i- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -i- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -i- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NRI SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: 007132; AAA61783.1;
 DR EMBL: 014534; AAA58594.1;
 DR HSSP: P03372; THCU.
 DR MIM: 600380;
 DR InterPro: IPR000536;
 DR InterPro: IPR001628;
 DR Pfam: PF00104; hormone_rec.1.
 DR Pfam: PF00105; z1-C4.1.
 DR PRINTS: PR00047; STROIDEINGER.
 DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KW Zinc-finger.
 FT DNA_BIND 87 154 NUCLEAR RECEPTOR-TYPE.
 FT ZN_BINDING 87 107 C4-TYPE.
 FT ZN_BINDING 125 149 C4-TYPE.
 FT DOMAIN 231 461 LIGAND-BINDING (POTENTIAL).
 FT DOMAIN 173 176 POLY-GIN.
 FT DOMAIN 191 196 POLY-SER.
 SO SEQUENCE 461 AA: 51102 MW: 6803309PBC50BE CRC64:

Query Match 21.6%; Score 472; DB 1; Length 461;
 Best Local Similarity 28.3%; Fred. No. 1.1e-30;
 Matches 122; Conservative 80; Mismatches 119; Indels 110; Gaps 14;

37 GPQICRGVCGDKATGTFHFNWTCGCKGFFRRAKRMARLRCPRKGCACETIKRT 95
 83 GHELCKRWGCDKASGRHYVULSCGCKGFFRRKSVYGAARVACRGCGTGGDAMRRKKQ 142
 96 ACRLLKCLSGMKKMTMSDAVERRALIRKKSEKGTGTO---PLGVG----- 142
 143 QCRLEKCKEAGMRQCVLSEDIKRRK--IKQDQESQSQSQSVGVQVQSSSSASGNGA 200
 143 -----LTFEJRMMLREIMDAQMKTFTDTFSHKFNRLPGVLSAGE 183
 201 SSGGSEASQSQSGEEGVQTLTAQELMLOLVAAOLOCNKKSF----- 244
 184 LPESIQARREERAKMSVKRLDLSLKVSIQREGDSVMYKPPADSGKEIFSLPHM 243
 245 -----DQKRVTPW-----LQAD-----PQSRARQGRFA--H 271
 244 ADMSTYMKRGLISPAKVISYFRDLPIEDQISLKGAFELCOLRENTVNAETGMEG 401
 272 TELATISVOEIVDFAKQVPGFLQREGDQIALKSTFEIMLETARVNHET--EC11 428
 302 ---GRISVLEED-TAGCFQOLLLEPMLEKPHYMKQJLHEEYVLMQALISFSPDPCVLO 358
 329 FLKDTYSKDFHRCAGLOVERINPIFFSKAMRRLGLDAEYALIALINISARPNVQE 388
 359 HRVVDQLOEPAITLKSYLECNRPQAHRL--LFLKIMAMLT-----EFA----- 401
 389 PGRVADQOPVEALISTRIKRPQDQLRPPEMLKIVSLKTLSSVHSEGVFALRLQKK 448
 402 -TPLMQLPLFGI 411
 449 LPLLSEIWDV 459

RESULT 15

Thu Jul 12 14:23:41 2001 }

us-09-276-935b-14.rsp

•
•
•
•

2

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 09:09:55 : Search time 16 Seconds

(without alignments)
1971.017 Million cell updates/sec

Title: US-09-276-935B-14

Perfect score: 2187
Sequence: 1 LEYRKESNNHADFVHCEDT.....AMLEPATPLMDELFGITGS 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922.5	42.2	386	2	thyroid hormone re
2	922.5	42.2	388	3	benzoate X recepto
3	771	35.3	420	2	vitamin D receptor
4	768.5	35.1	448	2	vitamin D receptor
5	762	34.8	423	2	1,25-dihydroxyvita
6	753.5	34.5	422	2	vitamin D receptor
7	750	34.3	427	2	vitamin D receptor
8	750	34.3	427	2	vitamin D receptor
9	748.5	34.2	425	2	vitamin D receptor
10	731	33.4	438	2	vitamin D receptor
11	472	21.6	446	2	steroid hormone-re
12	463	21.2	446	2	retinoid X recepto
13	461.5	21.1	445	2	steroid hormone re
14	461	21.1	446	2	orphan nuclear rec
15	456.5	20.9	447	2	nuclear orphan rec
16	426	19.5	369	1	thyroid hormone re
17	426	19.5	369	2	beta-thyroid hormo
18	425	19.4	373	2	gene c-erbA-beta p
19	424	19.4	373	2	thyroid hormone re
20	423	19.3	456	1	thyroid hormone re
21	419	19.2	373	2	thyroid hormone re
22	417	19.1	461	2	thyroid hormone re
23	417	19.1	461	2	thyroid hormone re
24	417	19.1	461	2	thyroid hormone re
25	413.5	18.9	878	2	ecdysone receptor
26	411.5	18.8	536	2	ecdysteroid recept
27	411	18.6	469	2	retinoid X-activa
28	406.5	18.6	484	2	retinoid X recepto
29	406	18.6	416	2	thyroid hormone re

ALIGNMENTS

Query Match	Best Local Similarity	Score	DB 2	Length	Matches	Conservative	Mismatches	Indels	Gaps	10:
18	EDTESVPKPSVNADEVGPOICRVCCKATGYHNYMTGCKGFFRRAMKRNARLRC	77	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
14	EEEDASNSCGIDEDGDPKICRACGDRATGYHFNAMTEGCKGFFRRAMKRNARLRC	73	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
78	PFRRKACETIKTRKQCACRLKCLIESGKKKKKMSDEAVERKALIKK-KSRHGTQ	136	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
74	PE-QNSVINSNRHHCACRLKCLDGMRKELMSDAVEORKALIKKRLIKLPP	132	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
137	PLVGGLLEEDGRMMRELMDAOMKTFOTTFHFNKRLPGVLSUGCELPELSAPSKREA	136	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
133	PPGA-SLTPFQGHPLTVLGAAHTKTFDPNFTFSKFR	171	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
197	AKMSVRRKDLCLAVSLDRLGEGDSVWYKPPADSGKEITSLPHMAOMKTYMKIIS	256	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
172	-----RSSDPT--GEQDPTS--SEATLMLPHISDLYTYMKITIS	207	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
257	FAKVSTYRDLPIEDQISLLKGAFFELCOLRPRTVFNATGTWGRGRLSTLELT-AGGF	315	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
208	FAKMLPYFKSLDIEDQIALKGSVAEVSVPFNFNFSNDITWECGPTPTDTEMLFAGF	267	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
316	QQLLEPLMKFHYMKLQLHEFEVYLMQATISPCDPEVGLQRRVYVLDQEGATILKS	375	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
268	ROLFLEPLVRIHRMKRNKLNLOSEETAMMAALSTRASDRPGVCMERKIQLOESHIALTKD	327	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
376	YIEECR-QOPARHRELFLTKAMLTFF-----ATPLMOELNG	410	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	
328	FIDSRPRSPQNRILATPKTMCCTITELRTVNDIHSKQMLETDNDIQNDATPLKRRVFG	383	thyroid hormone receptor homology - African clawed frog	386	194	54	99	69	10	

179 SSGCELPSTLQAPSRFAAKMSQVAKKLCSTLKVSUQLGEGDSVWYKPPADSGAKEIFS 238
 194 SODPSSEDSNVPKSDAGAFPEPMEPOMFSNIDISEDSSTPSMNIHLPH-----LP 246
 239 LLPHMAOMSTYMFKQIISPAVAVSYFRDLPEDQISTLKGAAPFLCQLRFTVNAETGT 298
 247 MLPPLADLVSYISQKIVGAFKMLPGEFRLDTADQIATLKSSAIEVIMLRMSWSTMEJMS 306
 299 WECGR--LSTYCLIED-TAGCFQULLLEPLKFKHYMLKQLQHEEYVLMQALISFSPDRPG 355
 307 WTGGSNDPKYKVSVDYQAGHSMDLPLPKFYQVGLKRLNHEEBVLLMAICILSPDRPG 366
 356 VLQHHVVDLOBOAFITKSYTECRRPQAPHRPLFLKIMAMLTFE----- 400
 367 VQDTSLVSTISDRISLDTQITIKCHHPPGSRLLYAKMIQIADLRSLNEESHKQYKCLS 426
 401 -----ATPLMOELFG 410
 427 FQPEHSMQTLPLVLEVPF 444

RESULT 5

A11761

1,25-dihydroxyvitamin D-3 receptor - rat
 C:Species: Rattus norvegicus (Norway 31)
 C>Date: 07-Jun-1990 *sequence_revision: 07-Jun-1990 *text_change 20-Sep-1999

C:Accession: A11761; A11367
 R:Bumester, J.K.; Maeda, N.; Deluca, H.F.

Proc. Natl. Acad. Sci. U.S.A. 85, 9499-9502, 1988

A:Title: Structure and regulation of the rat 1,25-dihydroxyvitamin D-3 receptor.
 A:Reference number: A11761; MUID:8071726

A:Accession: A11761

A:Molecule type: mRNA

A:Residues: 1-423 <BUB>
 A:Cross-references: GB:J04147; GR:J03630; NID:q203956; PIDN:AAA41089.1; PID:q203957

R:Bumester, J.K.; Maeda, N.; Deluca, H.F.

Proc. Natl. Acad. Sci. U.S.A. 85, 1005-1009, 1988

A:Title: Isolation and expression of rat 1,25-dihydroxyvitamin D-3 receptor cDNA.
 A:Reference number: A11367; MUID:86124963

A:Accession: A11367

A:Molecule type: mRNA

A:Residues: 1-382 <R02>
 C:Superfamily: unassigned erba-related proteins; erba transforming protein homology
 C:Keywords: DNA binding; transcription regulation; zinc finger
 F:22-337/Domain: erba transforming protein homology <ERBA>
 F:24-44/Region: zinc finger
 F:60-84/Region: zinc finger

Query Match 34.8%; Score 762; DR 2; Length 423;
 Best Local Similarity 40.0%; Pred. No. 2,9e-52;

Matches 164; Conservative 70; Mismatches 128; Indels 48; Gaps 8;

38 POLICVGCDAKATGYHFNVMTCGCGFFRRAMKRNALRCFFRKACETIKTRQOCAC 97
 21 PRICVGCDAKATGYHFNVMTCGCGFFRRAMKRNALRCFFRKACETIKTRQOCAC 79
 98 RLKRCLESGMKKEMIMSDAEVEERKALIKRKSERTGTQPLGVGGLTEEGRMIRELMDA 157
 80 RLKRCVIGMKKEFTLDEEYQKREIMKKEEFALKDSLRPK-LSPGQHIITALLDA 138
 158 QMKTFDTTFSEHFKNRLGVL--SSGCELPSTLQAPSRFAAKMSQVAKKLCSTLKVSUQLGEGDSVWYKPPADSGAKEIFS 238
 139 HHKTYDPTVADPFRDPRPVKMDSGSYSPR---PILSFGSNSSSSSDLYTSLDME 194
 212 -----SLDRCGDSVWYKPPADSGAKEIFSLLPHMAOMSTYMFKQIISPAVAVSYFRD 298
 195 PSQFSGNDLNGEDSD-----DPSVTLDSLSPSLMLPLADLVSYISQKIVGAFKMLPGEFRLDTADQIATLKSSAIEVIMLRMSWSTMEJMS 306
 267 LPIEDQISTLKGAAPFLCQLRFTVNAETGTWECGR--LSTYCLIED-TAGCFQULLLEPLKFKHYMLKQLQHEEYVLMQALISFSPDRPG 355
 250 LPSDQIATLKSSAIEVIMLRMSWSTMEJMSDQSYQVYKVDVDSKAGHTLELLEPL 309

324 LKEHYMKKIQIHEEYVLMQALISFSPDRPGVQHHVVDLOBOAFITKSYTECRRPQAPHRPLFLKIMAMLTFE----- 400
 310 IKFOVGLKRLNHEEBVLLMAICIVSPDRGVQDAKLVATQIRLSNLTQITIKCHHP 369
 384 FAHREPLKIMAMLTFE-----ATPLMOELFG 410
 370 PGSHQIYAKMIQIADLRSLNEESHKQYKCLSFPQENSMKLTPLVLEVPF 444

RESULT 6

PC4019

vitamin D receptor - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C>Date: 26-Jun-1995 *sequence_revision 14-Jul-1995 *text_change 20-Sep-1995

C:Accession: PC4019
 R:Kamei, Y.; Kawada, T.; Fukuwatari, T.; Ono, T.; Kato, S.; Sugimoto, E.

Gene 152, 281-282, 1995

A:Title: Cloning and sequencing of the gene encoding the mouse vitamin D receptor.
 A:Reference number: PC4019; MUID:95137405

A:Accession: PC4019

A:Molecule type: mRNA

A:Residues: 1-422 <RAM>
 A:Cross-references: DBR:031969; NID:q699618; PIDN:BA06737.1; PID:d1007311; PID:q499

C:Superfamily: unassigned erba-related proteins; erba transforming protein homology
 C:Keywords: DNA binding; zinc finger
 F:22-336/Domain: erba transforming protein homology <ERBA>
 F:24-89/Domain: DNA binding *status predicted <B1N>
 F:191-422/Region: vitamin D binding *status predicted

Query Match 34.5%; Score 753.5; DR 2; Length 422;
 Best Local Similarity 39.9%; Pred. No. 1.3e-51;

Matches 166; Conservative 69; Mismatches 120; Indels 61; Gaps 10;

38 POLICVGCDAKATGYHFNVMTCGCGFFRRAMKRNALRCFFRKACETIKTRQOCAC 97
 21 PRICVGCDAKATGYHFNVMTCGCGFFRRAMKRNALRCFFRKACETIKTRQOCAC 79
 98 RLKRCLESGMKKEMIMSDAEVEERKALIKRKSERTGTQPLGVGGLTEEGRMIRELMDA 157
 80 RLKRCVIGMKKEFTLDEEYQKREIMKKEEFALKDSLRPK-LSPGQHIITALLDA 138
 158 QMKTFDTTFSEHFKNRLGVL--GVLSGCELP-----PSIQAPSRFAAKMS 200
 139 HHKTYDPTVADPFRDPRPVKMDSGSYSPRPLSFGSNSSSDLYTSLDMEWASTS 198
 201 QVRKDLCSLKVSUQLGEGDSVWYKPPADSGAKEIFSLLPHMAOMSTYMFKQIISPAVAVSYFRD 298
 199 TM--DL-----NDEGS-----DPSVTLDSLSPSLMLPLADLVSYISQKIVGAFKMLPGEFRLDTADQIATLKSSAIEVIMLRMSWSTMEJMS 306
 261 LPIEDQISTLKGAAPFLCQLRFTVNAETGTWECGR--LSTYCLIED-TAGCFQULLLEPLKFKHYMLKQLQHEEYVLMQALISFSPDRPGVQHHVVDLOBOAFITKSYTECRRPQAPHRPLFLKIMAMLTFE----- 400
 243 LPIEDQISTLKGAAPFLCQLRFTVNAETGTWECGR--LSTYCLIED-TAGCFQULLLEPLKFKHYMLKQLQHEEYVLMQALISFSPDRPGVQHHVVDLOBOAFITKSYTECRRPQAPHRPLFLKIMAMLTFE----- 400
 318 LPIEDQISTLKGAAPFLCQLRFTVNAETGTWECGR--LSTYCLIED-TAGCFQULLLEPLKFKHYMLKQLQHEEYVLMQALISFSPDRPGVQHHVVDLOBOAFITKSYTECRRPQAPHRPLFLKIMAMLTFE----- 400
 303 ELIPPLIKFYQVGLKRLNHEEBVLLMAICIVSPDRGVQDAKLVATQIRLSNLTQITIKCHHP 369
 378 ECNRPQAPHRPLFLKIMAMLTFE-----ATPLMOELFG 410
 363 RCRHPPQSHQIYAKMIQIADLRSLNEESHKQYKCLSFPQENSMKLTPLVLEVPF 444

RESULT 7

A28200

vitamin D receptor - human
 C:Species: Homo sapiens (man)
 C>Date: 19-May-1989 *sequence_revision 19-May-1989 *text_change 20-Sep-1999

C:Accession: A28200; I55353
 R:Baker, A.R.; McDougal, D.P.; Hughes, M.; Crisp, T.M.; Mandelstam, D.J.; Haussler, J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3294-3298, 1988

A:Title: Cloning and expression of full-length cDNA encoding human vitamin D receptor.
 A:Reference number: A28200; MUID:86217887

5

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 09:09:55 : Search time 22.98 seconds

(without alignments)
1092.182 Million cell updates/sec

Title: US-09-276-935b-14

Perfect score: 2187
Sequence: 1 LEVPRKESNNHADFHCEDT.....AMLTFTATPLMGLGIGTGS 414

Scoring table: RUSUM62
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

A-Geneseq_0601.*
1: /SID58/qcadata/geneseq/geneseq/AA1980.DAT.*
2: /SID58/qcadata/geneseq/geneseq/AA1981.DAT.*
3: /SID58/qcadata/geneseq/geneseq/AA1982.DAT.*
4: /SID58/qcadata/geneseq/geneseq/AA1983.DAT.*
5: /SID58/qcadata/geneseq/geneseq/AA1984.DAT.*
6: /SID58/qcadata/geneseq/geneseq/AA1985.DAT.*
7: /SID58/qcadata/geneseq/geneseq/AA1986.DAT.*
8: /SID58/qcadata/geneseq/geneseq/AA1987.DAT.*
9: /SID58/qcadata/geneseq/geneseq/AA1988.DAT.*
10: /SID58/qcadata/geneseq/geneseq/AA1989.DAT.*
11: /SID58/qcadata/geneseq/geneseq/AA1990.DAT.*
12: /SID58/qcadata/geneseq/geneseq/AA1991.DAT.*
13: /SID58/qcadata/geneseq/geneseq/AA1992.DAT.*
14: /SID58/qcadata/geneseq/geneseq/AA1993.DAT.*
15: /SID58/qcadata/geneseq/geneseq/AA1994.DAT.*
16: /SID58/qcadata/geneseq/geneseq/AA1995.DAT.*
17: /SID58/qcadata/geneseq/geneseq/AA1996.DAT.*
18: /SID58/qcadata/geneseq/geneseq/AA1997.DAT.*
19: /SID58/qcadata/geneseq/geneseq/AA1998.DAT.*
20: /SID58/qcadata/geneseq/geneseq/AA1999.DAT.*
21: /SID58/qcadata/geneseq/geneseq/AA2000.DAT.*
22: /SID58/qcadata/geneseq/geneseq/AA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2187	100.0	414	20	AAV42691 Human pregnane X r
2	2167	99.1	434	20	AAV15931 A human intranucle
3	2167	99.1	457	20	AAV16035 A human intranucle
4	2167	99.1	457	20	AAV15932 A human intranucle
5	2167	99.1	466	20	AAV25410 Human nmr7 partial
6	2167	99.1	473	20	AAV25411 Human nmr7-1 prote
7	2167	99.1	473	20	AAV15936 A human intranucle
8	2167	99.1	473	20	AAV15933 A human intranucle
9	2167	99.1	473	20	AAV09516 Human vitamin D re
10	2165	99.0	437	20	AAV09515 Human vitamin D re
11	2049	93.7	434	20	AAV21799 Human steroid and

12	1463	66.9	316	20	AAV42689
13	910.5	41.6	386	17	AAV98521
14	762	34.8	423	20	AAV94623
15	759	34.7	423	19	AAV47509
16	750	34.3	427	19	AAV68156
17	750	34.3	427	20	AAV09064
18	750	34.3	450	20	AAV09064
19	750	34.3	477	20	AAV09035
20	731	33.4	348	18	AAV32536
21	731	33.4	348	18	AAV3902
22	726	33.2	348	14	AAV41346
23	688.5	31.5	357	20	AAV17872
24	680	31.1	358	20	AAV93903
25	656	30.0	356	19	AAV37261
26	656	30.0	356	20	AAV94622
27	593	27.1	367	14	AAV43656
28	472	21.6	460	16	AAV74738
29	472	21.6	461	15	AAV52980
30	472	21.6	461	17	AAV97982
31	472	21.6	461	17	AAV98140
32	470	21.5	460	18	AAV25034
33	466.5	21.3	445	21	AAV32374
34	465	21.3	461	17	AAV6234
35	464	21.2	446	17	AAV9736
36	461	21.1	446	17	AAV94169
37	457	20.9	443	16	AAV74739
38	457	20.9	443	18	AAV25035
39	456.5	20.9	447	17	AAV03326
40	446.5	20.4	440	14	AAV3744
41	428.5	19.6	757	20	AAV28603
42	428.5	19.6	757	22	AAV67096
43	426.5	19.5	757	19	AAV71297
44	426	19.5	456	9	AAV80921
45	424.5	19.4	472	19	AAV40072

ALIGNMENTS

RESULT 1
ID AAV42691 standard: Protein: 414 AA.
AC AAV42691:
XX
XX
XX 17-JAN-2000 (first entry)
XX
XX
XX Human pregnane X receptor (hpxr).
XX
XX
XX Human: nuclear receptor: pregnane X receptor: PKC: CYP: CYP3A4:
XX
XX cytochrome P-450 mono-oxygenase: drug interaction: hpxr.
XX
XX Homo sapiens.
XX
XX
XX W09948915-A1.
XX
XX 30-SEP-1999.
XX
XX 26-MAR-1999: 99MO-US06737.
XX
XX 27-MAR-1998: 98US-0079593.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX KILMER SA, WILLSON TM;
XX
XX WPI: 1999-601202/51.
XX
XX N-PSDB: AAZ07997.
XX
XX New human pregnane X receptor, used to identify specific modulators and
XX agents that induce expression of cytochrome P-450 mono-oxygenase
XX
XX Claim 4; Fig 1A-D; 69pp; English.

Human pregnane X re
Xenopus orphan rec
Kat vitamin D rece
Kat vitamin D rece
Human vitamin D re
Human vitamin D re
Human vitamin D re
Human vitamin D re
Constitutively act
Human CAR receptor
Human CAR receptor
Mouse nuclear rece
Mouse CAR receptor
Kat vitamin D rece
Kat vitamin D rece
Human ubiquitons n
Human ubiquitons n
Human recombinant
Human steroid rece
NPR receptor pote
Human ubiquitons n
Mouse CNRP-1, Mu
Human foetal lung
Retinoid X recepto
OR-1 orphan recept
Kat ubiquitons nuc
Kat ubiquitons nuc
LXR-alpha, orhan
XR2, Homo sapiens
ECR polypeptide su
L. cuprina cytochrome
Junilia cuprina cy
Sequence encoded b
Human retinoid rec

XX The invention provides an isolated human nuclear receptor (designated
 CC program: x receptor, pxx) that binds to a cyclochrome p 450 mono-oxygenase
 CC (CYP) promoter. The hpxR is used to identify its specific modulators,
 CC and compounds that induce CYP9A4 expression (i.e., to identify drug
 CC interactions, since CYP9A4 is involved in many biotransformations of
 CC drugs). The modulators are particularly useful for associating particular
 CC diseases and conditions with pxx and for treating such conditions.
 CC Antisense raised against hpxR can be used for decimation and
 CC purification of hpxR. The present sequence represents the hpxR.

XX
 CC Sequence: 414 AA:

Query Match 100.0%; Score 2187; DB 20; Length 414;
 best local similarity 100.0%; Pred. No. 4, 26-209;
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LEVPRKESWNIADVFHETPSVPGKSVNADDEVGCHQICRVCGRKATGYHFNMTCEG 60
 DB 1 LEVPRKESWNIADVFHETPSVPGKSVNADDEVGCHQICRVCGRKATGYHFNMTCEG 60
 QY 61 CKGFPRKAMKRNARLRCGPRKGCACETTRKTRGQCACRLKCLKESGKKKEMINSDEAVEE 120
 DB 61 CKGFPRKAMKRNARLRCGPRKGCACETTRKTRGQCACRLKCLKESGKKKEMINSDEAVEE 120
 QY 121 RKALIKKKSEBRTGTOPICVGGITTEGQMMIRELMDQMKTFPTTSHKRNFRPGVLS 180
 DB 121 RKALIKKKSEBRTGTOPICVGGITTEGQMMIRELMDQMKTFPTTSHKRNFRPGVLS 180
 QY 181 GCELPESIQAPSRFAKWSQVRKDLCSLKVSQILKQEDGSVNNYKPPADSGKEIFSL 240
 DB 181 GCELPESIQAPSRFAKWSQVRKDLCSLKVSQILKQEDGSVNNYKPPADSGKEIFSL 240
 QY 241 PHMAIMSTYMKKGISFAKVISYFROLPLEDOISLKGAFELCOLRNFVNAETGWE 300
 DB 241 PHMAIMSTYMKKGISFAKVISYFROLPLEDOISLKGAFELCOLRNFVNAETGWE 300
 QY 360 CGRLSYCTEDTAGGFOQLLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLE 360
 DB 360 CGRLSYCTEDTAGGFOQLLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLE 360
 QY 401 VVDQLODFATILKSYECNRPOFARHFLFLKIMAMTEETVTPIMQELFGITGS 414
 DB 401 VVDQLODFATILKSYECNRPOFARHFLFLKIMAMTEETVTPIMQELFGITGS 414

RESULT 2

AAV15941 standard; protein: 414 AA.

XX AAV15941:

XX 04 AUG 1999 (first entry)

DE A human intranuclear receptor protein.

KM Human; intranuclear receptor protein; drug development; diagnosis;

KW treatment.

OS Homo sapiens.

PN JP1112787.2 A.

PD 18 MAY 1999.

PE 07 AUG 1998; 98JP-0224172.

PN 11 AUG 1997; 97JP-0240435.

PA (NISHI) JAPAN PHARMACO INC.

PK WPI: 1999-450430/80.

DE N-PSDB: AAK5966.

XX New intranuclear receptor protein - useful for drug development and

PT diagnosis and treatment of disease

XX claim 1; Page 15-16; 48pp; Japanese.

CC The present sequence represents a human intranuclear receptor protein.

CC The nucleic acid sequence was isolated from a human adult cDNA

CC library using a swedish AMO23 derived probe. The protein can

CC be used for the development of drugs and diagnosis and treatment

CC of various diseases.

XX
 CC Sequence: 414 AA:

Query Match 99.1%; Score 2167; DB 20; Length 414;
 best local similarity 95.4%; Pred. No. 4, 36-207;
 Matches 414; Conservative 0; Mismatches 0; Indels 20; Gaps 1;

QY 1 LEVPRKESWNIADVFHETPSVPGKSVNADDEVGCHQICRVCGRKATGYHFNMTCEG 60
 DB 1 LEVPRKESWNIADVFHETPSVPGKSVNADDEVGCHQICRVCGRKATGYHFNMTCEG 60
 QY 61 CKGFPRKAMKRNARLRCGPRKGCACETTRKTRGQCACRLKCLKESGKKKEMINSDEAVEE 120
 DB 61 CKGFPRKAMKRNARLRCGPRKGCACETTRKTRGQCACRLKCLKESGKKKEMINSDEAVEE 120
 QY 121 RKALIKKKSEBRTGTOPICVGGITTEGQMMIRELMDQMKTFPTTSHKRNFRPGVLS 180
 DB 121 RKALIKKKSEBRTGTOPICVGGITTEGQMMIRELMDQMKTFPTTSHKRNFRPGVLS 180
 QY 181 GCELPESIQAPSRFAKWSQVRKDLCSLKVSQILKQEDGSVNNYKPPADSGKEIFSL 240
 DB 181 GCELPESIQAPSRFAKWSQVRKDLCSLKVSQILKQEDGSVNNYKPPADSGKEIFSL 240
 QY 241 PHMAIMSTYMKKGISFAKVISYFROLPLEDOISLKGAFELCOLRNFVNAETGWE 300
 DB 241 PHMAIMSTYMKKGISFAKVISYFROLPLEDOISLKGAFELCOLRNFVNAETGWE 300
 QY 360 CGRLSYCTEDTAGGFOQLLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLE 360
 DB 360 CGRLSYCTEDTAGGFOQLLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLEPMLE 360
 QY 401 VVDQLODFATILKSYECNRPOFARHFLFLKIMAMTEETVTPIMQELFGITGS 414
 DB 401 VVDQLODFATILKSYECNRPOFARHFLFLKIMAMTEETVTPIMQELFGITGS 414

RESULT 3

AAV16045 standard; protein: 457 AA.

XX AAV16045:

XX 04 AUG 1999 (first entry)

DE A human intranuclear receptor protein.

KM Human; intranuclear receptor protein; drug development; diagnosis;

KW treatment.

OS Homo sapiens.

PN JP1112787.2-A.

PD 18 MAY 1999.

PE 07 AUG 1998; 98JP-0224172.

QY		CGRLSYCEDIAGCFOQLLEPMLKEHYMKKKLOLHEEYVMAOISLSPDRPCVLQHR	360
Dd			399
QY		361 VVDOLDFQFATLEKSYECNRPQAHREPLFKIMAMLTET-----F	400
Dd			459
OY		400 vvdqfgeqfatllkkyecmripghrfrllikimamltelsinaqhqlrrllidghpf	
OY		401 ATPLMGELEFGITGS 414	
Dd			
Dd		460 atplmgelilgits 473	
RESULT	7		
AAV15936	ID	AAV15936 standard; Protein: 473 AA.	
XX			
AC		AAV15936;	
XX			
DT		04-AUG-1999 (first entry)	
XX			
DE		A human intranuclear receptor protein.	
XX			
KW		Human; intranuclear receptor protein; drug development; diagnosis;	
XX		treatment.	
XX			
OS		Homo sapiens.	
PN		JP11127872-A.	
PD		18-MAY-1999.	
XX			
PF		07-AUG-1998; 98JP-0224172.	
XX			
PR		11-AUG-1997; 97JP-0230335.	
XX			
PA		(NISB) JAPAN TOBACCO INC.	
DR		WPI: 1999-350330/30.	
N-PSDB:		AAK59974.	
PT		New intranuclear receptor protein - useful for drug development and	
PT		diagnosis and treatment of disease	
PS			
XX		Disclosure: Page 32-35; 38pp; Japanese.	
CC		The present sequence represents a human intranuclear receptor protein.	
CC		The nucleic acid sequence was isolated from a human adult cDNA	
CC		library using a swedish ANO23 derived probe. The protein can	
CC		be used for the development of drugs and diagnostics and treatment	
CC		of various diseases.	
XX			
S0		Sequence 473 AA:	
Query Match	99.1%; Score 2167; DB 20; Length 473;		
Best Local Similarity	95.4%; Pred. No. 4.9e+207;		
Matches 414; Conservative	0; Mismatches 0; Indels 20; Gaps		1;
OY		1 LEVRKESNNHADYVHCEDTESVPCKPSYNADDEVGCPQICRVCDDKATGYHNVNTCEG	60
Dd			
Dd		40 Ievrpkesnhdafvhcetesvpkpsynadeevygpqifrcvcdkalgyhinvmctcg	99
OY		61 CGGFRRBAKNNAARLCRCPRKGACETTRKTRROGCACRLKCIEGMKEIMSDAEVE	120
Dd			
OY		121 RRALKIRKKSEKTGPQLCVQCLTEFORMMIREIMDAQMKTFDTPFSHKNFRLPGVLS	180
Dd			
Dd		160 rralikrkksertlqplqvglteegrmrirelmdaqmkftdtlshknrlpgvlss	219
OY		181 GGELFESLGAPSRERFAAKKSQYVRKDLCSIKVSLQJRGEDGSWMYTKPPADSGKELFILL	240

Db	220	gcelpeslqapsreedaakmsvrvklcslkxslqrgcdqsvwnykpdpadsqkqetlsl	279
Qy	241	PHADMSITVMKGIISFKAVISYFRDPIEDQISLKGAAETCOLRNIYVNAFTGME	300
Db	280	phmadmslymfkqjlsstakvisyfrdpliedqislkgaateqlrntvnaetqrwe	339
Qy	301	CGRTSYCEDTAGGFOOLLEPMLEKFHYMLKKLIQHEEYVIMQALSLSPDRKVLQHR	360
Db	340	cgrtsyciedtagdqrgqlllepmklfhymlkqlgheceyvlmqaislfsprdrvgqtr	409
Qy	361	VVDOLQEDPAITLKSYLEPCNRQPAHREPLTKINAMLT-----F	400
Db	400	vvdqlqegfatllkylecnpqpahrefllklnamlelrslnaqhtqlrlrhdhpf	459
Qy	401	ATPLMQLFGITGS 414	
Db	460	atplmqelrftqgs 473	
RESULT 8			
Db	AAV15933	AAV15933 standard; Protein: 473 AA.	
AC	AAV15933:		
XX			
DI	04-AUG-1999	(first entry)	
XX			
DE	A human intranuclear receptor protein.		
XX			
KW	Human: intranuclear receptor protein; drug development; diagnosis;		
XX	treatment.		
OS	Homo sapiens.		
XX			
FN	JP11127872-A.		
XX			
PD	18-MAY-1999.		
XX			
FP	07-AUG-1998;	98JP-022417Z.	
XX			
PR	11-AUG-1997;	97JP-0230335.	
XX			
PA	(NLSB) JAPAN TOBACCO INC.		
XX			
DR	WPI: 1999-350330/30.		
DR	N-PSDB: AAX59968.		
XX			
PT	New intranuclear receptor protein - useful for drug development and		
XX	diagnosis and treatment of disease		
PS	Claim 2: Page 17-19: 38pp: Japanese.		
XX			
CC	The present sequence represents a human intranuclear receptor protein.		
XX	The nucleic acid sequence was isolated from a human adult cDNA		
CC	library using a swelffish ANO23 derived probe. The protein can		
CC	be used for the development of drugs and diagnosis and treatment		
CC	of various diseases.		
XX			
SQ	Sequence 473 AA:		
Qy	Query Match	99.1% Score 2167; DH 20; Length 473;	
Db	Best Local Similarity 95.4%;	Pred. No. 4.9e-207;	
	Matches 414; Conservative 0; Mismatches 0; Indels 29; Gaps		
Qy	1 LEVREKESNNADFWHCEDTESVPKPSVNADEWGGQICRWGSDKATGYHNVMTCEG	60	
Db	40 levrkesnnadfvhcedtesvpkpsvnadeevgpbqicrvcdkatqymhnmvltceq	99	
Qy	61 CGGFRBRAAKRNALRLCKPRKGCACETTRKTRGCGACGRKRLKLGSKMKEMTMSIAVFE	120	
Db	100 cggftrbraakrnatrlcpkrlgacelttrtrrgqacitrlklesqmkkemtmsleavce	159	

0y	121	RRALKKKKSSRTTQPLAVGTLPEEOMMMRLRLMDQOMTETOTESHENFKRPEVLSS	180
10b	160	LTLLKKKKSSRTTQPLAVGTLPEEOMMMRLRLMDQOMTETOTESHENFKRPEVLSS	219
0y	181	GTPLPEPSLOASREKFAAKMSVYKKDLSTLKVSTLOLRKEEDSSVMNRYKRPALSTAKFPLSL	240
10b	220	QQLPEPSLOASREKFAAKMSVYKKDLSTLKVSTLOLRKEEDSSVMNRYKRPALSTAKFPLSL	279
0y	241	PRMALDMSYMKRGLSPAKVTSYKRDLELEQSLSTLKGAAFLQALPNTVFNAETGWE	300
10b	280	PRMALDMSYMKRGLSPAKVTSYKRDLELEQSLSTLKGAAFLQALPNTVFNAETGWE	339
0y	301	GRGLSYCTLEPFAACEPQGLLEPRMKKPLMYMKRLQHEEYVMQALSLSPDRDGYQGR	360
10b	340	GRGLSYCTLEPFAACEPQGLLEPRMKKPLMYMKRLQHEEYVMQALSLSPDRDGYQGR	399
0y	361	VWLDQDPGPAITLKSYLETNDNQHAIHEFLPLKIMADTE	400
10b	400	VWLDQDPGPAITLKSYLETNDNQHAIHEFLPLKIMADTE	459
0y	401	ATPLMDELPTGGS	414
10b	460	ATPLMDELPTGGS	474

RESUME	9
AAV09516	AAV09516 standard; protein; 47% AA.
XX	
AC	AAV09516;
XX	
UT	16 JUL 1999 (first entry)
XX	
DE	Human vitamin D receptor related gamma 2 protein.
XX	
KW	Human; vitamin D receptor related protein; VDR; obesity; diabetes;
KW	anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;
KW	hypercholesterolaemia; hyperlipoproteinaemia; osteoporosis; tumour;
KW	hyperproliferative skin disorder; hyperthyroidism.
OS	
DS	Ibmo Sapiens.
XX	
PN	W0919454-A1.
XX	
PD	22 APR 1999.
XX	
FE	31 AUG 1998; 98WO SE0154R.
FR	SI MAR 1998; 98SE-00014R.
PR	14 OCT 1997; 97SE-00047A5.
XX	
PA	(PINA) PHARMACIA & UPJOHN AB.
XX	
F1	Forrestam A., Dahlborg M.
XX	
OR	WP: 1999 802508/25.
DE	N PDBH: AAX56241.
XX	
FT	New vitamin D receptor related (VDR) polypeptides, useful for
PT	treating obesity, diabetes, anorexia and rheumatoid arthritis
XX	
IS	Claime [9; Page 22-24; 4pp; English.
XX	
CV	The present sequence is a human vitamin D receptor related (VDR)
CV	polypeptide. Human VDR polypeptides and substances which affect VDR
CV	signal transduction, can be used for treating metabolic, proliferative
CV	or inflammatory conditions. They can be used in the manufacture of a
CV	medicament for treating the following conditions: obesity, diabetes,
CV	anorexia, lipoprotein defects, hyperlipidemia, hypercholesterolemia or
CV	hyperlipoproteinaemia and osteoporosis, rheumatoid arthritis, benign and
CV	malign tumours, hyperproliferative skin disorders or hyperthyroidism.
CV	Nucleotide acid vectors encoding for expression of a VDR polypeptide can

XX Sequence 473 AA:
 50
 55
 60
 65
 70
 75
 80
 85
 90
 95
 100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995
 1000

Query Match	99.18;	Score	2167;	DB	20;	Length	474;
Best Local Similarity	95.48;	Pred. No.	4, 9e-207;				
Matches	414;	Conservative	0;	Mismatches	0;	Indels	20;
						Gaps	1

CY	1	LEVERKUSENINADVHEREDTESVGKISVNNDREVCCHQICRNGCBKALCYHNEVMTEFG	60
Dib	40	IovrppkssmhadtVhrvcdLscvpkjksvnadveevqqrptetvqqdkaTqyhtvmvtceq	99
CY	61	GKGFEERAAKKBNARLCDFPERGACETTRKGRROGCALFKLKCLPSCOKKEMDSDAVEE	140
Dib	100	CKqftrfrmkfnarLrcpLtrkaacelLrkrttgyacrLrkclcsnmkkmsinsdevoee	189
CY	121	KKALLKRKKSRERTGTPLGVGLTERBORMRLBELMDQMKTFOTTHSHFNRPILVLLSS	180
Dib	160	rtaLtkkrksortatcpIqvqqlLcevgpmnrlcundaqmklidLtslkoDlpvylss	219
CY	181	GTELPESTIOASREERAAKMSOVRRKLCSLKAYSLDRJHDGDSVMNYKPAPASSCKELPSL	240
Dib	220	qccelpeslqpasrecaakswsqvrdLcskvsldqLredqsavwvykpapassckelLslsll	279
CY	241	PBMADRSVTMRKGIITSPAKVLSTFRDLPTFEOJLSLKGAAFLHCULPHNTVFNAETGWE	300
Dib	280	pImadrslymrkaiIslakvisyrrdlplcdqslLkqaalelcqLIntvlncetqtwe	349
CY	301	CGRLSYCTLETPAGCEFOULLEPMLKPYMYLKKLOLHEBEVLMQATLSNSDRIVYGHK	360
Dib	400	cgrlsyctlecfagactqpllLcommlktrymlkkloqhnecvyylmqalslspdrprvqghr	399
CY	401	VWDLQCEYFATILEKSYLPCTNRPOVAHHCELEKIMAMITE.....F	400
Dib	430	vwdlqcyfatllksytocnrppahrltlLklnamltcelrstnchgtgllLqghlhpL	459
CY	401	AALPMOPELPGTTGS_414	
Dib	460	alpmoPeLatlpgs_474	

RESIDU	16	
AA09515		
10	AA09515, standard; protein; 437 AA.	
XX		
AC	AA09515;	
XX		
D1	16-JUN, 1999 (first entry)	
XX		
DE	Human vitamin D receptor related gamma protein.	
XX		
KW	Human: vitamin D receptor related protein; VDR; obesity; diabetes;	
KW	diarrhea; rheumatoid arthritis; lipoprotein defect; hyperlipidemia	
KW	hypercholesterolemia; hyperlipoproteinemia; osteoporosis; tumour;	
KW	hyperproliferative skin disorder; hyperhydriosis.	
XX		
OS	Homo sapiens.	
PN	W091935A-A1.	
XX		
PD	22 APR 1999.	
XX		
PF	41-ADG-1998; 9880-SKU1548.	
XX		

10 AAW47509 standard; Protein: 423 AA.
 XX
 AC AAW47509;
 XX
 DE 26 JUN 1998 (first entry)
 XX
 DE rat vitamin D receptor (VDR).
 XX
 KM rat vitamin D receptor; isoform protein; VDR1; VDR0; diagnostics;
 KM dominant negative receptor; signal transduction channel;
 KM bone density disorder; screening.
 XX
 OS Rattus rattus.
 XX
 FH Key: location/qualifiers
 FT Misc difference: 360
 FT /note: "encoded by cDNA"
 XX
 XX W0947172 A1.
 XX
 PD 18 DEC 1997.
 XX
 PF 10 JUN 1997; 97W0-180947.
 XX
 PR 10 JUN 1996; 96JP-0194179.
 XX
 PA (CHUS) CHUGAI PHARM CO LTD.
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Kato S, Ueno K;
 XX
 DR W01: 1998-051917/05.
 DR N PDB: AAW04129.
 XX
 PT DNA encoding a vitamin D receptor isoform protein - useful for bone
 PT density determination and for screening substances for vitamin D
 PT activity
 XX
 PS Disclosure: Fig 1; 46pp; Japanese.
 XX
 CC A novel cDNA sequence encodes the rat vitamin D receptor isoform
 CC protein (VDR). The isoform differs from the normal receptor
 CC (VDR0), which comprises the present sequence, in having the
 CC vitamin D response element curtailed by 66 residues, and having an
 CC extra 19 residues inserted at the C terminal of this element. It
 CC acts as a dominant negative receptor in the vitamin D signal
 CC transmission channel.
 CC The isoform protein can be used to diagnose bone density disorders,
 CC and screen for substances having potential vitamin D like activity.
 XX
 SO Sequence 423 AA;
 Query Match 34.78; Score 759; DB 19; Length 423;
 Best Local Similarity 40.08; Pred. No. 5, 50-67;
 Matches 164; Conservative 69; Mismatches 129; Indels 48; Gaps 8;
 QY 48 PULCVCNKKATCYEENVTETCKCFPRKRNKAKQCEPRKQACEETKTRQCAV 97
 1 11 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 21 ptiqyqatratqthnmtceqekottrtsukktlittcpf-mndettkkntthcpac 79
 1
 QY 98 RLRKTLSSQMKKEMPSDAVEERKALRKQKSEKRTGYPICVGNLDPFGRRMRLMDA 157
 1
 DB 80 rikrvdimmkclttdcevpkrominkrocvtkdskltpk-lseegphltdltda 138
 1
 QY 158 QMKTEPTTSSEKNEPLPVL--SSKTELPESLQAPSRKRAKRSQVRRDLSIKV- 211
 1
 DB 139 hkttyqpyadlrdlrrpymdqtqsyvpr - pldstsssssssdlytstidmo 194
 1
 QY 212 ----SLQKQKSGSWNWKPKPADSSTKLESLHFMALSTYMEKQLISFAKVLSTYPRD 266
 1
 DB 195 pssltstldlmpdsd-----qlpsvtdlspismpladlvssyqkvrtakmipqtrd 249

QY 267 LFTNDQSLKGAAPFLQALPNTVFNARETCWEGRLSYCTAEPT--ANQFQALTEPM 323
 1 11 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 256 ltsdqlvltlksaalevlnltsngstlmdmswdvsgdykvdvtdvskadhltell 309
 1
 QY 424 LRFHYMLKKLQDHEEYVLMQALSLSPDRGVQGHKVVQLOPQALTKSTYERMRQ 483
 1
 DB 410 lktqvdkklnlhoovhlnatctvspdrpovqdklvoadtrdrtsrldtytrcthpp 469
 1
 QY 484 PAHREPLFKIMAMLTET-----ATPMQELFEG 413
 1
 DB 470 pssltstldlmpdsd-----qlpsvtdlspismpladlvssyqkvrtakmipqtrd 249

Search completed: July 11, 2001, 09:10:27
 Job time: 42 sec

Thu Jul 12 14:23:38 2001

us-09-276-935b-14.rag

Page 11

2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 09:10:50 : Search time 24.22 Seconds

(without alignments)
2261.531 Million cell updates/sec

Title: US-09-276-935b-14

Perfect score: 2187
Sequence: 1 LEVPRKESNNHAFVHCEDT.....AMLEPAPLMOELLFCITGS 414

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mbc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP unclassified: *
13: SP vertebrate: *
14: SP virus: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2167	99.1	473	4 Q9UNW4	Q9UNW4 homo sapien
2	2158	98.7	457	4 Q9UJ26	Q9UJ26 homo sapien
3	2156	98.6	434	4 Q9UJ27	Q9UJ27 homo sapien
4	1938.5	88.6	420	4 Q9UJ23	Q9UJ23 homo sapien
5	1936.5	88.5	397	4 Q9UJ24	Q9UJ24 homo sapien
6	1851	84.6	379	4 Q9UJ25	Q9UJ25 homo sapien
7	1700.5	77.8	411	6 Q9TU02	Q9TU02 cyctolagus
8	1631.5	74.6	342	4 Q9UJ22	Q9UJ22 homo sapien
9	922.5	42.2	386	13 Q91B39	Q91B39 xenopus lae
10	922.5	42.2	388	13 Q9DF24	Q9DF24 xenopus lae
11	876.5	40.1	391	13 Q9DM43	Q9DM43 gallus galli
12	771	35.3	420	13 Q91B73	Q91B73 parallachth
13	748.5	34.2	425	13 Q91B74	Q91B74 parallachth
14	746	34.1	453	13 Q9PTM2	Q9PTM2 brachydanio
15	446	20.4	397	5 Q76246	Q76246 uca pugnati
16	434.5	19.9	784	5 Q9GPH1	Q9GPH1 calliphora
17	433	19.8	386	13 Q9PVE4	Q9PVE4 brachydanio
18	431	19.7	673	5 Q76827	Q76827 ceratitis c
19	428.5	19.6	680	5 Q9U3Y4	Q9U3Y4 aedes albop

20	424.5	19.4	472	4 Q92943	Q92943 homo sapien
21	423	19.3	416	13 Q9W6N4	Q9W6N4 hippodossu
22	423	19.3	456	4 Q13966	Q13966 homo sapien
23	413.5	18.9	881	5 Q9V9K8	Q9V9K8 drosophila
24	411	18.8	469	11 Q62735	Q62735 rattus norv
25	410.5	18.8	541	5 Q97095	Q97095 locusta mig
26	406.5	18.6	484	11 Q60641	Q60641 mus musculu
27	402.5	18.4	491	5 Q02035	Q02035 temblorio mo
28	401	18.3	454	13 Q9W524	Q9W524 tuva tubrip
29	395.5	18.1	401	13 Q9U382	Q9U382 calina mos
30	393	18.0	416	13 Q9W785	Q9W785 salmo salar
31	393	18.0	447	13 Q9W623	Q9W623 tuva tubrip
32	390.5	17.9	457	13 Q90272	Q90272 brachydanio
33	387.5	17.7	444	5 Q44336	Q44336 amblyomma a
34	387	17.7	560	5 Q44337	Q44337 amblyomma a
35	387	17.7	570	5 Q44338	Q44338 amblyomma a
36	386	17.6	444	13 Q90271	Q90271 brachydanio
37	386	17.6	444	13 Q91391	Q91391 brachydanio
38	385.5	17.6	513	5 Q77240	Q77240 choristoneu
39	385.5	17.6	541	5 Q77255	Q77255 choristoneu
40	384	17.6	455	13 Q9W6B3	Q9W6B3 coturnix co
41	383	17.5	459	11 Q9QWJ1	Q9QWJ1 rattus norv
42	382.5	17.5	452	13 Q91155	Q91155 notophthalm
43	381	17.4	367	13 Q90273	Q90273 brachydanio
44	381	17.4	448	13 Q92019	Q92019 xenopus lae
45	381	17.4	555	14 Q96594	Q96594 avian eryth

ALIGNMENTS

```
RESULT 1
Q9UNW4 ID Q9UNW4 PRELIMINARY; PRT: 473 AA.
AC Q9UNW4:
DT 01-MAY-2000 (TRENKLEL 13, created)
DT 01-MAY-2000 (TRENKLEL 13, last sequence update)
DT 01-MAR-2001 (TRENKLEL 16, last annotation update)
DE ORPHAN NUCLEAR RECEPTOR.
GN PAR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=98445350; PubMed=9770465;
RA Bertilsson G., Heidrich J., Svensson K., Asman M., Jendeborg L.,
RA Sydow-Backman M., Ohlsson R., Postlund H., Blomquist P.,
RA Berkenstam A.;
RT Identification of a human nuclear receptor defines a new signaling
RT pathway for CYP3A induction.*
RL Proc. Natl. Acad. Sci. U.S.A. 95:12208-12213(1998).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
DR EMBL: AF084644; AAC64557.1; -.
DR HSSP: P10826; IHRB.
DR InterPro: IPR000536; -.
DR InterPro: IPR001628; -.
DR InterPro: IPR001723; -.
DR Pfam: PF00104; hormone_rec.1.
DR PRINTS: PF00105; zf-c4.1.
DR PRINTS: PR00398; STRDHOMNER.
DR PRINTS: PR00447; STROIDFINGER.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR.1.
DR SMART: SM00399; Znf_C4.1.
DR DNA-binding: Nuclear protein; Receptor; Transcription regulation;
KM Zinc-finger.
SQ
SEQUENCE 473 AA: 53899 MW: 3049023G9B903016 CRC64;
Query Match 99.1%; Score 2167; DB 4; Length 473;
```


Best Local Similarity 95.48; Pr: 1 No. 4,66-178;
 Match: 414; Conserved: 0; Mismatches: 0; Indels: 20; Gaps: 1;

UY 1 LEVRRKSSNNADVHCHETESVGRKPSVNADEFGVQICVQVQDKATGCHFNMTGEG 60
 |||||
 DB 40 LEVRRKSSNNADVHCHETESVGRKPSVNADEFGVQICVQVQDKATGCHFNMTGEG 99

UY 61 CKFFRRKRNARLKCFFRRKACETTRKTRGCVARLKRKLKESKKKEMIMSEAAVEE 120
 |||||
 DB 100 CKFFRRKRNARLKCFFRRKACETTRKTRGCVARLKRKLKESKKKEMIMSEAAVEE 159

UY 121 KRALIKRRKSRKTOTQPLAVGGLTEEDQNNMIRELMQAMKTPDTTFSEKFNRIQVLS 180
 |||||
 DB 160 KRALIKRRKSRKTOTQPLAVGGLTEEDQNNMIRELMQAMKTPDTTFSEKFNRIQVLS 219

UY 181 KRALIKRRKSRKTOTQPLAVGGLTEEDQNNMIRELMQAMKTPDTTFSEKFNRIQVLS 240
 |||||
 DB 220 GCELPESLQAVSRREAAKWSQVRKDLCSLKVSLQKCHDSVWVYKPPADSGCKEIPSL 279

UY 241 PHMAIMSTYMKGLISPAKVISYPRDLPEQDLSLKGAAPFLQVLRNIVFAETGWE 300
 |||||
 DB 280 PHMAIMSTYMKGLISPAKVISYPRDLPEQDLSLKGAAPFLQVLRNIVFAETGWE 339

UY 301 CKGRSVCTLEDIAGPQOALLPEMLKPFHYMLKKGQIHEEYVIMQALSLSPDRIGVQHR 360
 |||||
 DB 340 CKGRSVCTLEDIAGPQOALLPEMLKPFHYMLKKGQIHEEYVIMQALSLSPDRIGVQHR 399

UY 361 VVLOLOPQFATLTKSYLTCNRPQAHREFLKIMAMLETE -----F 400
 |||||
 DB 400 VVLOLOPQFATLTKSYLTCNRPQAHREFLKIMAMLETELSINAGTQRLKIDIHPE 459

UY 401 ATPLMDELPGTGS 414
 |||||
 DB 460 ATPLMDELPGTGS 473

RESULT 2
 ID Q90J26 PRELIMINARY: PRT: 457 AA.
 AC Q90J26;
 DT 01 MAY 2000 (TERMINOT: 14, Created)
 DT 01 MAY 2000 (TERMINOT: 14, Last sequence update)
 DT 01 MAR 2001 (TERMINOT: 16, Last annotation update)
 DE NUCLEAR HORMONE RECEPTOR PRK1 C.
 GN PER1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID: 9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE: LIVER;
 RA Heard D.J., Holloway J., Hansen C., Tommerup N., Aagaard L.,
 RA Vissing H.;
 RT Identification of a novel protein isoform of the human nuclear
 RT hormone receptor PRK/SKX and localization to chromosome 3q12.1
 RT 14,377;
 RL Eur. J. Hum. Genet. 0:0-0(0).
 CC 1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC 1 SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
 EMU: A1009946; CAH5490.1;
 DB HSSP: P10826; IIRA.
 DB InterPro: IPR000546;
 DB InterPro: IPR001628;
 DB InterPro: IPR001723;
 DB Pfam: PF00104; hormone_rcv_1.
 DB Pfam: PF00105; Z1_C4; 1.
 DB PRINTS: PR00498; STRKRDHOMNER.
 DB PRINTS: PR00047; STRKRDHOMNER.
 DB PRINTS: PS00041; NUCLEAR_RECEPTOR; 1.
 DB SMART: SM00499; Znf_C4; 1.
 KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
 KW Zinc finger.

NO SEQUENCE 457 AA; 52145 MW; 812FBH354DH744E GRCH43;
 Query Match 98.78; Score 2158; DB 4; Length 457;
 Best Local Similarity 94.98; Pr: 0; Mismatches: 2; Indels: 20; Gaps: 1;
 Match: 414; Conserved: 0; Mismatches: 0; Indels: 20; Gaps: 1;

UY 1 LEVRRKSSNNADVHCHETESVGRKPSVNADEFGVQICVQVQDKATGCHFNMTGEG 60
 |||||
 DB 24 LEVRRKSSNNADVHCHETESVGRKPSVNADEFGVQICVQVQDKATGCHFNMTGEG 83

UY 61 CKFFRRKRNARLKCFFRRKACETTRKTRGCVARLKRKLKESKKKEMIMSEAAVEE 120
 |||||
 DB 84 CKFFRRKRNARLKCFFRRKACETTRKTRGCVARLKRKLKESKKKEMIMSEAAVEE 143

UY 121 KRALIKRRKSRKTOTQPLAVGGLTEEDQNNMIRELMQAMKTPDTTFSEKFNRIQVLS 180
 |||||
 DB 144 KRALIKRRKSRKTOTQPLAVGGLTEEDQNNMIRELMQAMKTPDTTFSEKFNRIQVLS 203

UY 181 GCELPESLQAVSRREAAKWSQVRKDLCSLKVSLQKCHDSVWVYKPPADSGCKEIPSL 240
 |||||
 DB 204 GCELPESLQAVSRREAAKWSQVRKDLCSLKVSLQKCHDSVWVYKPPADSGCKEIPSL 263

UY 241 PHMAIMSTYMKGLISPAKVISYPRDLPEQDLSLKGAAPFLQVLRNIVFAETGWE 300
 |||||
 DB 264 PHMAIMSTYMKGLISPAKVISYPRDLPEQDLSLKGAAPFLQVLRNIVFAETGWE 323

UY 301 CKGRSVCTLEDIAGPQOALLPEMLKPFHYMLKKGQIHEEYVIMQALSLSPDRIGVQHR 360
 |||||
 DB 324 CKGRSVCTLEDIAGPQOALLPEMLKPFHYMLKKGQIHEEYVIMQALSLSPDRIGVQHR 383

UY 361 VVLOLOPQFATLTKSYLTCNRPQAHREFLKIMAMLETE -----F 400
 |||||
 DB 384 VVLOLOPQFATLTKSYLTCNRPQAHREFLKIMAMLETELSINAGTQRLKIDIHPE 443

UY 401 ATPLMDELPGTGS 414
 |||||
 DB 444 ATPLMDELPGTGS 457

RESULT 4
 ID Q90J27 PRELIMINARY: PRT: 414 AA.
 AC Q90J27;
 DT 01 MAY 2000 (TERMINOT: 14, Created)
 DT 01 MAY 2000 (TERMINOT: 14, Last sequence update)
 DT 01 MAR 2001 (TERMINOT: 16, Last annotation update)
 DE NUCLEAR HORMONE RECEPTOR PRK1 A.
 GN PER1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID: 9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE: LIVER;
 RA Heard D.J., Holloway J., Hansen C., Tommerup N., Aagaard L.,
 RA Vissing H.;
 RT Identification of a novel protein isoform of the human nuclear
 RT hormone receptor PRK/SKX and localization to chromosome 3q12.1
 RT 14,377;
 RL Eur. J. Hum. Genet. 0:0-0(0).
 CC 1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC 1 SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
 EMU: A1009946; CAH5489.1;
 DB HSSP: P10826; IIRA.
 DB InterPro: IPR000546;
 DB InterPro: IPR001628;
 DB InterPro: IPR001723;
 DB Pfam: PF00104; hormone_rcv_1.
 DB Pfam: PF00105; Z1_C4; 1.
 DB PRINTS: PR00498; STRKRDHOMNER.
 DB PRINTS: PR00047; STRKRDHOMNER.

DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
DR SMART: SM00399; ZNF_C4; 1.
KW DNA-binding: Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 434 AA; 49777 MW; AB36DE02304C4200 CRC64;

Query Match 98.6%; Score 2156; DB 4; Length 434;
Best Local Similarity 94.7%; Pred. No. 2,8e-177;
Matches 411; Conservative 1; Mismatches 2; Indels 20; Gaps 1;

```
QY 1 LEVPRKESWNHADVHCEDETSVPKPSVNADEVGGPQICVCGDKATGYHFNMTCEG 60
DB 1 MEVRKESWNHIDVHCEDTESVPKPSVNADEVGGPQICVCGDKATGYHFNMTCEG 60
QY 61 CKGFRRAMKRNARLRCFPRKGACETTRKTRQCAQRLKCLSGMKKEMMSDAVEE 120
DB 61 CKGFRRAMKRNARLRCFPRKGACETTRKTRQCAQRLKCLSGMKKEMMSDAVEE 120
QY 121 RRALIKRKKSERGTQPLGVGGLTEEDORMMIRELMDQMKTFDTTFSHKFNRPLGVLS 180
DB 121 RRALIKRKKSERGTQPLGVGGLTEEDORMMIRELMDQMKTFDTTFSHKFNRPLGVLS 180
QY 181 GCELPESLQAPSRREBAKMSQVRKDCSLKVSLOLKGEDGSVMYKPPADSGKEIFSL 240
DB 181 GCELPESLQAPSRREBAKMSQVRKDCSLKVSLOLKGEDGSVMYKPPADSGKEIFSL 240
QY 241 PHMADMSTYMKGIISPAKVISYFRDLPLEDOISLKGAEFLCOLPNTVFNATGTWE 300
DB 241 PHMADMSTYMKGIISPAKVISYFRDLPLEDOISLKGAEFLCOLPNTVFNATGTWE 300
QY 301 CGRLSTCYLEDTAGGFOQLLEPMLKFHYMLKKLOLHEEYVLMQATSLFSPDRGVLOHR 360
DB 301 CGRLSTCYLEDTAGGFOQLLEPMLKFHYMLKKLOLHEEYVLMQATSLFSPDRGVLOHR 360
QY 361 VVDLOEQFATILKSYIECNRPQPAHREPLKIMAMTE-----F 400
DB 361 VVDLOEQFATILKSYIECNRPQPAHREPLKIMAMTE-----F 400
QY 401 ATPLMQELFGITGS 414
DB 401 ATPLMQELFGITGS 434
```

RESULT 4
ID Q90J23 PRELIMINARY; PRT; 420 AA.
AC Q90J23;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE NUCLEAR HORMONE RECEPTOR PRR2-C.
CN PRR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Heard D.J., Holloway J., Hansen C., Tommerup N., Aagaard L.,
RA Vissing H.;
RT *Identification of a novel protein isoform of the human nuclear
RT hormone receptor PKR/SXR and localization to chromosome 3q12.1
RT -13.3.*;
RL Eur. J. Hum. Genet. 0:0-0(0).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
EMBL: AJ009937; CAB55493.1; -.
DR HSSP: P10826; IHRA.
DR InterPro: IPR000536; -.
DR InterPro: IPR001628; -.
DR InterPro: IPR001723; -.

DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; Z1-C4; 1.
DR PRINTS: PR00398; STROHORMONR.
DR PRINTS: PR00447; STROIDFINGER.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
DR SMART: SM00399; ZNF_C4; 1.
KW DNA-binding: Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 420 AA; 48149 MW; CF4D9E1256859165 CRC64;

Query Match 88.6%; Score 1938.5; DB 4; Length 420;
Best Local Similarity 86.4%; Pred. No. 1.4e-158;
Matches 375; Conservative 0; Mismatches 2; Indels 57; Gaps 2;

```
QY 1 LEVPRKESWNHADVHCEDETSVPKPSVNADEVGGPQICVCGDKATGYHFNMTCEG 60
DB 1 LEVPRKESWNHIDVHCEDTESVPKPSVNADEVGGPQICVCGDKATGYHFNMTCEG 63
QY 24 LEVPRKESWNHIDVHCEDTESVPKPSVNADEVGGPQICVCGDKATGYHFNMTCEG 83
DB 24 LEVPRKESWNHIDVHCEDTESVPKPSVNADEVGGPQICVCGDKATGYHFNMTCEG 83
QY 61 CKGFRRAMKRNARLRCFPRKGACETTRKTRQCAQRLKCLSGMKKEMMSDAVEE 120
DB 61 CKGFRRAMKRNARLRCFPRKGACETTRKTRQCAQRLKCLSGMKKEMMSDAVEE 120
QY 84 CKGFRRAMKRNARLRCFPRKGACETTRKTRQCAQRLKCLSGMKKEMMSDAVEE 143
DB 84 CKGFRRAMKRNARLRCFPRKGACETTRKTRQCAQRLKCLSGMKKEMMSDAVEE 143
QY 121 RRALIKRKKSERGTQPLGVGGLTEEDORMMIRELMDQMKTFDTTFSHKFNRPLGVLS 180
DB 121 RRALIKRKKSERGTQPLGVGGLTEEDORMMIRELMDQMKTFDTTFSHKFNRPLGVLS 180
QY 144 RRALIKRKKSERGTQPLGVGGLTEEDORMMIRELMDQMKTFDTTFSHKFNRPLGVLS 196
DB 144 RRALIKRKKSERGTQPLGVGGLTEEDORMMIRELMDQMKTFDTTFSHKFNRPLGVLS 196
QY 181 GCELPESLQAPSRREBAKMSQVRKDCSLKVSLOLKGEDGSVMYKPPADSGKEIFSL 240
DB 181 GCELPESLQAPSRREBAKMSQVRKDCSLKVSLOLKGEDGSVMYKPPADSGKEIFSL 240
QY 197 -----VSLQRLKRGDGSVMYKPPADSGKEIFSL 226
DB 197 -----VSLQRLKRGDGSVMYKPPADSGKEIFSL 226
QY 241 PHMADMSTYMKGIISPAKVISYFRDLPLEDOISLKGAEFLCOLPNTVFNATGTWE 300
DB 241 PHMADMSTYMKGIISPAKVISYFRDLPLEDOISLKGAEFLCOLPNTVFNATGTWE 300
QY 227 PHMADMSTYMKGIISPAKVISYFRDLPLEDOISLKGAEFLCOLPNTVFNATGTWE 286
DB 227 PHMADMSTYMKGIISPAKVISYFRDLPLEDOISLKGAEFLCOLPNTVFNATGTWE 286
QY 301 CGRLSTCYLEDTAGGFOQLLEPMLKFHYMLKKLOLHEEYVLMQATSLFSPDRGVLOHR 360
DB 301 CGRLSTCYLEDTAGGFOQLLEPMLKFHYMLKKLOLHEEYVLMQATSLFSPDRGVLOHR 360
QY 287 CGRLSTCYLEDTAGGFOQLLEPMLKFHYMLKKLOLHEEYVLMQATSLFSPDRGVLOHR 346
DB 287 CGRLSTCYLEDTAGGFOQLLEPMLKFHYMLKKLOLHEEYVLMQATSLFSPDRGVLOHR 346
QY 361 VVDLOEQFATILKSYIECNRPQPAHREPLKIMAMTE-----F 400
DB 361 VVDLOEQFATILKSYIECNRPQPAHREPLKIMAMTE-----F 400
QY 401 ATPLMQELFGITGS 414
DB 401 ATPLMQELFGITGS 420
```

RESULT 5
ID Q90J24 PRELIMINARY; PRT; 397 AA.
AC Q90J24;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE NUCLEAR HORMONE RECEPTOR PRR2-A.
CN PRR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Heard D.J., Holloway J., Hansen C., Tommerup N., Aagaard L.,
RA Vissing H.;
RT *Identification of a novel protein isoform of the human nuclear
RT hormone receptor PKR/SXR and localization to chromosome 3q12.1
RT -13.3.*;
RL Eur. J. Hum. Genet. 0:0-0(0).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
EMBL: AJ009937; CAB55492.1; -.

KA Smith D.P., Mason C.S., Jones E.A., and R.M.W.
 RT "A novel nuclear receptor superfamily member in *Xenopus* that
 associated with RXR, and shares extensive sequence similarity to the
 mammalian vitamin D3 receptor."
 RL Nucleic Acids Res. 22:6671(1994).
 CC 1. SUBCELLULAR LOCATION: NUCLEAR (by SIMILARITY).
 CC 2. SIMILARITY TO CA-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
 DR EMBL: X75163; CAAS006.1; .
 DR HSSP: P10826; HBA.
 DR InterPro: IPR000536; .
 DR InterPro: IPR001628; .
 DR Pfam: PF00104; hormone rec. 1.
 DR Pfam: PF00105; Zf-C4_1.
 DR PRINTS: PR00047; STEROID RECEPTOR.
 DR PROSITE: PS00041; NUCLEAR RECEPTOR; UNKNOWN_1.
 DR SMART: SM00430; H04.1.
 DR DNA binding: Nuclear protein; receptor; transcription regulation;
 KM Zinc finger.
 SO SEQUENCE: 486 AA; 44302 MW; 880463219556574 Cn664.

Query Match 42.2% Score 922.5; DB 13; Length 380;
 Best Local Similarity 46.6% Pred. No. 2,76-71;
 Matches 194; Conservative 54; Mismatches 99; Indels 69; Gaps 10;

UY 18 EFTESVYKPKSVNADEVGQVQICVGVGKATGYHFNMTGCGKGFPRKRNARLAC 77
 DB 14 EEEELANSNGTGEDEEDKEDPKICACGCRATGYHFNMTGCGKGFPRKRNARLAC 73
 UY 78 PERKACETIKTRKROVQACRLKCTESGKMKEMLSDAVEERKALIKRKSKECTG 136
 DB 74 PF-QNSGVINNNRBRHCACRLKCTESGKMKEMLSDAVEERKALIKRKSKECTG 132
 UY 147 PLAVGCLTEHQRMMIRELMADQKPTPTTSHKNEIRLGVLSGCTELPSLGASREFA 196
 DB 143 PEGV-STLPEQCHETELVEAHKTEPNEPFSKNER-----PIK-- 171
 UY 197 AKMSQVRKRLSLKVSQJRGESVNNYKPPADSCKETPSLPHMAKSTYMKGLIS 256
 DB 172 -----KSDPT---QDPQATIS-SEATLMLPHISLPTMYMKGLIS 207
 UY 257 FAKVSYRDLPLEQSLKGAPELQLRNIVNAETWEGTGRSLSTLEP-AGEP 315
 DB 208 FAKMLPYRSLALEQALIKGSVAEYVTRNTVNSGTINIRNPPRYTDTMLAPR 267
 UY 416 GCLLEPMKLFHYMKKQJHEEYVLMQALISFSDRGVYLQHVYVQJQVQATLTKS 375
 DB 268 RQLEPLVYRIHRRMKRLNLSSEYVMMALSLFASDRGVYDMEKLTQKQELATLTKD 427
 UY 476 YLEPNR LQPAHRLPLKIMAMTEF-----ATLMQPLNG 410
 DB 428 FLDSDPSPQNDLKYKIMELTELRLVNDIHSKQLLEIMQJQVATPLMRVFG 484

RESULT 10
 G99F24 PRELIMINARY: PRT: 488 AA.
 ID G99F24:
 AC G99F24:
 DT 01-MAR-2001 (TREMbled: 16, Cleaved)
 DT 01-MAR-2001 (TREMbled: 16, last sequence update)
 DT 01-MAR-2001 (TREMbled: 16, last annotation update)
 DE ORIGIN: NUCLEAR RECEPTOR BXR-BEFA.
 OS *Xenopus laevis* (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
 OC Xenocephalia; *Xenopus*.
 CX NCBI_TaxID 8355;
 RN 111
 RP SEQUENCE FROM N.A.
 RA Nishikawa J., I., Saito K., Sasaki M., Tanihara Y., Nishihara T.,
 RT "Molecular cloning and functional characterization of a novel nuclear
 RT receptor, stimulating embryonic benzoyl receptor BXR #2;
 RL Hirohara, H.ophys. Res. Commun. 0:0-0(2000).

DR EMBL: AF052013; AAG24910.1; .
 KM Receptor.
 SO SEQUENCE: 488 AA; 44512 MW; 76CA692E19BEE08 Cn664;

Query Match 42.2% Score 922.5; DB 13; Length 380;
 Best Local Similarity 47.1% Pred. No. 2,76-71;
 Matches 196; Conservative 49; Mismatches 102; Indels 69; Gaps 10;

UY 18 EFTESVYKPKSVNADEVGQVQICVGVGKATGYHFNMTGCGKGFPRKRNARLAC 77
 DB 15 EEEELANSNGTGEDEEDKEDPKICACGCRATGYHFNMTGCGKGFPRKRNARLAC 74
 UY 78 PERKACETIKTRKROVQACRLKCTESGKMKEMLSDAVEERKALIKRKSKECTG 136
 DB 75 PF-QNSGVINNNRBRHCACRLKCTESGKMKEMLSDAVEERKALIKRKSKECTG 133
 UY 147 PLAVGCLTEHQRMMIRELMADQKPTPTTSHKNEIRLGVLSGCTELPSLGASREFA 196
 DB 144 PEGV-STLPEQCHETELVEAHKTEPNEPFSKNER-----PIK-- 172
 UY 197 AKMSQVRKRLSLKVSQJRGESVNNYKPPADSCKETPSLPHMAKSTYMKGLIS 256
 DB 173 -----KSDPT---QDPQATIS-SEATLMLPHISLPTMYMKGLIS 208
 UY 257 FAKVSYRDLPLEQSLKGAPELQLRNIVNAETWEGTGRSLSTLEP-AGEP 315
 DB 209 FAKMLPYRSLALEQALIKGSVAEYVTRNTVNSGTINIRNPPRYTDTMLAPR 268
 UY 416 GCLLEPMKLFHYMKKQJHEEYVLMQALISFSDRGVYLQHVYVQJQVQATLTKS 375
 DB 269 RQLEPLVYRIHRRMKRLNLSSEYVMMALSLFASDRGVYDMEKLTQKQELATLTKD 428
 UY 476 YLEPNR LQPAHRLPLKIMAMTEF-----ATLMQPLNG 410
 DB 429 FLDSDPSPQNDLKYKIMELTELRLVNDIHSKQLLEIMQJQVATPLMRVFG 484

RESULT 11
 G99F24 PRELIMINARY: PRT: 491 AA.
 ID G99F24:
 AC G99F24:
 DT 01-MAR-2001 (TREMbled: 16, Cleaved)
 DT 01-MAR-2001 (TREMbled: 16, last sequence update)
 DT 01-MAR-2001 (TREMbled: 16, last annotation update)
 DE XENOBOTIC RECEPTOR.
 OS *Gallus gallus* (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archaeoptera; Aves; Mesobatrachia; Galliformes; Phasianidae;
 OC Gallus.
 CX NCBI_TaxID 9031;
 RN 111
 RP SEQUENCE FROM N.A.
 RA TISSUE: LIVER, KIDNEY, AND INTESTINE;
 RA PubMed 11005656;
 RA Handschin C., Fedivne M., Meyer U.A.;
 RT "cXKR, a chicken xenobiotic-sensing orphan nuclear receptor, is related
 RT to both mammalian pregnane X receptor (PXR) and constitutive
 RT androstane receptor (CAR)."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:10769-10774(2000).
 DR EMBL: AF276753; AAC18474.1; .
 KM Receptor.
 SO SEQUENCE: 491 AA; 43945 MW; 9A2636C9ZALAF106 Cn664;

Query Match 40.1% Score 876.5; DB 13; Length 391;
 Best Local Similarity 46.9% Pred. No. 2,56-67;
 Matches 188; Conservative 67; Mismatches 93; Indels 54; Gaps 13;

UY 15 DIESVYKPKSVNADEVGQVQICVGVGKATGYHFNMTGCGKGFPRKRNARLAC 73
 DB 8 DIESVYKPKSVNADEVGQVQICVGVGKATGYHFNMTGCGKGFPRKRNARLAC 63


```

0Y 74 RLCPKPKGAELTRKTRBOOQALREKCKJESGKKKMIIMSDUAVEHRAVL-1KKKKSS 132
Db 64 HFTCFTFR-SCPTIKARPOOQALOKCLDVGKKKMIISSEALGRKRLRJÖRMLAO- 121
QY 133 TGTOPLGVGQTEEOGRMIRELMAQMKTEFTTFSHKNR-----LPVLSGGCELP- 186
Db 122 --ADP---GGLTAQOQELSTILIAHKRTFDSSTSOYHOPAVKLCTPQCS---QSP- 173
QY 187 SLQAFREAKWSORRDKLSKVSJQKGGDSVWYKPRADSGKELFSLPHMAD 246
Db 174 GPQVPS-----ASLSPQDCLDEYVL-----PDVFSLPHAD 207
QY 247 STYHKKGLISAKEYSTFRDPLPIDOJLSLKGAAPFELQJLRENTVNAEFTGTECGRLST 306
Db 208 STFMIOQVIFPAKEKLPFERGRLPIDOJLSLKGLATGICQJQFNTVNEETNAMEGCGHCE 267
QY 307 CLEDA--GGPQOLLLEPMKRFYMLKXIOJHEPEYVIMQALISFSPRPREVJQHRVQDL 365
Db 268 TIKDQALAGFOOJYLEPLKRFISLKLRLHEATEVYLLVAMILESPDHASVTDQRFIDOL 327
QY 366 QGFQATILKSYIEGNCRFQPAHFETLFIIMAMILEFATPILQ 406
Db 328 QEKVALTIKSYIDHRHPRPDEGRFLYAKLLLLLLLELQJLKNHE 368

```

RESULT	12		
091B73			
ID	091B73	PRELIMINARY:	PRF: 420 AA.
AC	091B73:		
DT	01-OCT-2000 (TReMBLrel, 15, Created)		
DT	01-OCT-2000 (TReMBLrel, 15, Last sequence update)		
DE	01-MAR-2001 (TReMBLrel, 16, Last annotation update)		
CN	VITAMIN D RECEPTOR A.		
OS	VDR.		
OC	Paralichthys olivaceus (Flounder).		
OC	Euarthra; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorphi; Pleuronectiformes;		
OC	Pleuronectiformes; Pleuronectiformes; Pleuronectiformes;		
OX	NCBI_TaxID=8255;		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=INTESTINE;		
RC	MEDLINE=20200092; PubMed=10733902;		
RA	Suzuki T., Suzuki N., Srivastava A.S., Kurekawa T.;		
RT	*Identification of cDNAs encoding two subtypes of vitamin D receptor		
RL	in flounder, Paralichthys olivaceus.;		
RL	Biochem. Biophys. Res. Commun. 270:40-45(2000).		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
CC	-1- SIMILARITY: TO C4 TYPE STEROID RECEPTOR ZINC FINGER FAMILY.		
CC	EMBL: AB037674; BAA95016.1; -		
DR	InterPro: IPR000536; -		
DR	InterPro: IPR001628; -		
DR	InterPro: IPR001723; -		
DR	Pfam: PR00104; hormone_rec; 1.		
DR	Pfam: PR00105; zf-c4; 1.		
DR	PRINTS: PR00398; STRDOMKMER.		
DR	PRINTS: PR00047; STRDOMKMER.		
DR	PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.		
DR	SMART: SM00430; HOL1; 1.		
KW	DNA-binding; Nuclear protein; Receptor; Transcription regulation;		
KW	Zinc-finger.		
SO	SEQUENCE 420 AA: 47486 MW: 038F8F00D4F38067 CRC64:		

```

Query: March 35, 38; Score 771; DB 13; length 420;
      Best Local Similarity 41.0%; Pred. No. 3.2e-58;
      Matches 170; Conservative 66; Mismatches 137; Indels 42; Gaps 11
CY 29 VNAD: EVGSGPOLICRGVGGKATGATGHEENVNTGTCGGCGEFGFRKAKRNARLRCPPKAGCEIT 87
      - - - - - 1 - - - - - 1 - - - - - 1 - - - - - 1 - - - - - 1 - - - - - 1 - - - - -
DB 11 VSPDIEDNRAPIRGCGCGKATGCFHFNATGCGCKGCFHFRSKRKAISPCP-NGSCITTT 69

```

QY	88	RKIRROQOACRIRKOTLEGOMKKKIMSDENAEVBKRALIKRKRSTRTGTUPLUGVGLTFEO	147
Db	70	KONRHCQACRLKRCRIDGMRKEFTIDIEVQJRKEMIKRKEFEARERMRP-INEO	128
QY	148	KMMIRELMDAOMKTDTEFSHKKNBRLP-GLVSSOCELPESIOAPSRKBAKMSOVK	204
Db	129	ARMISLSLVEAHKITYDASIPDSFRPRPVRCGPATRSASRAKSLHSISDMSINSFINSP	188
QY	205	DLCSLKV-SHOLKGEQSWYMKPPADSGKELFSLIPMADMSYMKGLISFAKVI	261
Db	189	SV-DTKMFSNLIIMWYOGA-SSPSSEFENKLSMLPHALAVASVIGUVAFKMI	244
QY	262	STHROLPEDOSILKGAFFCOLURENTVINAETGWEG-PLSTGLD-TAGVGOOL	318
Db	244	PGFRRLMEDOIALIKSSAIEILIMRSKNSFLSDMSNSGGPFCXINDV KAGHILE	303
QY	319	LLEPMLKHHYMKKLOLHEEYVLMQALSLESPBRPVLOHRVNDGOLQOPAILLKAYIE	378
Db	304	LLEPVLKRVQVGLKRLNLEEBHVLLMGCLLSPRGVQVQAHAEVOLQIKPEALQAVIR	463
QY	379	CNRPOARFLEFLKIMATLTF-----ATPIMQELP	410
Db	364	IN-HPGORLLAKKIOKLADIRSLNEHSKOYSLSVQPHSMQITPLILEVNG	416

ID	Q91B74	PRELIMINARY:	PT:	425 AA.
AC	Q91B74:			
DT	01-OCT-2000 (TREMblrel, 15, Created)			
DT	01-OCT-2000 (TREMblrel, 15, last sequence update)			
DE	01-MAR-2001 (TREMblrel, 16, last annotation update)			
DE	VITAMIN D RECEPTOR R.			
CN	VDRB.			
OS	Paralichthys olivaceus (Flounder).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Acanthopterygii; Percomorphi; Pleuronectiformes;			
OC	Pleuronectoidae; Bothidae; Paralichthys.			
OX	NCHI_TaxID=8255;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=INTESTINE;			
RX	MEDLINE=20200092; PubMed=10733902;			
RA	Suzuki T., Suzuki N., Srivastava A.S., Kurokawa T.;			
RT	*Identification of cDNAs encoding two subtypes of vitamin D receptor			
RT	in flounder, Paralichthys olivaceus.*			
RL	Biochem. Biophys. Res. Commun. 270:40-45(2000).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
CC	-1- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.			
CC	EMBL: AB037673; BA095015.1; -			
DR	InterPro: IPR005336; -			
DR	InterPro: IPR001628; -			
DR	InterPro: IPR001723; -			
DR	Pfam: PF00104; hormone_rec: 1.			
DR	Pfam: PF00105; zf-C4: 1.			
DR	PRINTS: PR003498; STRDOCKMER.			
DR	PRINTS: PR000447; STRDPTNGER.			
DR	PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.			
DR	SMART: SM00430; HO1; 1.			
KW	DNA-binding; Nuclear protein; Receptor; Transcription regulation;			
KW	Zinc-finger.			
SQ	SEQUENCE 425 AA: 48174 MW: 781049CFAC265659 CRC64;			

	Query Match Similarity	34.2%	Score 748.5,	DB 13,	Length 425;
	Best Local Similarity	39.6%:	Pred. No. 2.8e-56;		
	Matches 171; Conservative	73;	Mismatches 133;	Indels	Gaps 12.
OY	22 SVPGKPSVADP-EVGSDQICHWGSKATGYHHNVTCETCKGGFFRRKKRNARLCTDFR	80			
	I L D : I I I : I				
Dd	4 TVYSSTLSIAEDFDNRMPIICCGCKGRKTGFTHNNMTCGGCAGFPRSSKRRAFTFLCPD	62			

[illegible]

KM	DNA binding	Nuclear protein; Receptor; Transcription regulation
Zinc	Zinc finger	
FT	NON-TER	
FT	VARIANT	ADP1SP → VAP1SP
SD	SEQUENCE	197 AA; 45467 MW; 996441212H5C55AB; CXC04;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2001, 09:09:55 : Search time 12.21 Seconds
(without alignments)
683.037 Million cell updates/sec

Title: US-09-276-935B-14

Perfect score: 2187

Sequence: 1 LEYRKESNNHADFVHCEDT.....AMLETFPLMDELFGITGS 414

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cqn2_6/ptodata/2/1aa/5A.COMB.pep:*

2: /cqn2_6/ptodata/2/1aa/5H.COMB.pep:*

3: /cqn2_6/ptodata/2/1aa/6A.COMB.pep:*

4: /cqn2_6/ptodata/2/1aa/6B.COMB.pep:*

5: /cqn2_6/ptodata/2/1aa/6C.COMB.pep:*

6: /cqn2_6/ptodata/2/1aa/Backlist1.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	750	34.3	427	4	US-08-764-870-11 Sequence 11, Appl
2	731	33.4	348	1	US-08-459-489-10 Sequence 10, Appl
3	731	33.4	348	1	US-08-458-686-10 Sequence 10, Appl
4	731	33.4	348	1	US-07-843-350C-10 Sequence 10, Appl
5	590	27.0	367	1	US-07-737-736B-4 Sequence 2, Appl
6	472	21.6	461	1	US-08-330-518-2 Sequence 2, Appl
7	472	21.6	461	1	US-08-330-283-2 Sequence 2, Appl
8	472	21.6	461	1	US-08-646-248-2 Sequence 2, Appl
9	472	21.6	461	1	PCT-US95-13924-2 Sequence 2, Appl
10	472	21.6	461	5	PCT-US95-13931-2 Sequence 2, Appl
11	470	21.5	460	1	US-08-342-411A-2 Sequence 2, Appl
12	464	21.2	446	2	US-08-372-652-3 Sequence 3, Appl
13	464	21.2	446	5	PCT-US95-16311-3 Sequence 3, Appl
14	457	20.9	443	1	US-08-342-411A-4 Sequence 8, Appl
15	456.5	20.9	440	1	US-08-333-358-8 Sequence 8, Appl
16	456.5	20.9	440	1	US-08-463-694-8 Sequence 8, Appl
17	456.5	20.9	440	1	US-08-694-501-8 Sequence 8, Appl
18	456.5	20.9	447	1	US-08-373-945-1 Sequence 2, Appl
19	424.5	19.4	472	1	US-08-496-631-2 Sequence 2, Appl
20	423	19.3	461	4	US-08-764-870-3 Sequence 3, Appl
21	410	18.7	355	6	5223606-4 Patent No. 5223606
22	409	18.7	469	3	US-08-372-183-2 Sequence 2, Appl
23	409	18.7	469	4	US-09-468-721-2 Sequence 2, Appl
24	409	18.7	469	5	PCT-US95-17023-2 Sequence 2, Appl
25	406.5	18.6	484	2	US-08-372-652-1 Sequence 1, Appl
26	406.5	18.6	484	5	PCT-US95-16311-1 Sequence 1, Appl
27	405.5	18.5	746	4	US-09-144-759-18 Sequence 18, Appl

ALIGNMENTS

28	405.5	18.5	764	4	US-09-144-759-20	Sequence 20, Appl
29	399.5	18.3	451	2	US-08-372-652-2	Sequence 2, Appl
30	399.5	18.3	451	5	PCT-US95-16311-2	Sequence 2, Appl
31	394	18.0	410	4	US-08-764-870-2	Sequence 2, Appl
32	394	18.0	410	6	5438126-2	Patent No. 5438126
33	390	17.8	410	4	US-08-764-870-1	Sequence 1, Appl
34	387	17.7	448	6	5223606-2	Patent No. 5223606
35	379.5	17.4	433	2	US-08-466-120-2	Sequence 2, Appl
36	379.5	17.4	433	5	PCT-US94-07266-2	Sequence 2, Appl
37	372	17.0	462	2	US-08-592-383-2	Sequence 2, Appl
38	372	17.0	462	2	US-08-592-383-2	Sequence 2, Appl
39	372	17.0	462	5	PCT-US92-07320A-4	Sequence 4, Appl
40	372	17.0	457	2	US-08-095-728B-4	Sequence 2, Appl
41	372	17.0	797	2	US-08-095-728B-2	Sequence 2, Appl
42	370	16.9	416	4	US-08-764-870-4	Sequence 4, Appl
43	368.5	16.8	403	2	US-08-592-383-4	Sequence 4, Appl
44	368	16.8	368	6	5223606-3	Patent No. 5223606
45	363.5	16.6	454	4	US-08-764-870-5	Sequence 5, Appl

RESULT 1
US-08-764-870-11
Sequence 11, Application US/08764870
Patent No. 6236946
GENERAL INFORMATION:
APPLICANT: Scanlan, Thomas S
APPLICANT: Baxter, John D
APPLICANT: Fletcher, Robert J
APPLICANT: Wagner, Richard L
APPLICANT: Kushner, Peter J
APPLICANT: Apriletti, James W
APPLICANT: West, Brian
TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
TITLE OF INVENTION: Binding Domains
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward
STREET: Five Palo Alto Square, 3000 El Camino Real
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,870
FILING DATE: 13-DEC-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,540
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,543
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,606
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Jackie N
REGISTRATION NUMBER: 35,966
REFERENCE/DOCKET NUMBER: UCAL-246/0105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)843-5000
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 427 amino acids
TYPE: amino acid

Thu Jul 12 14:23:39 2001

us-09-276-935b-14.rai

Page 2

STRANDEDNESS: 1
TOPOLOGY: linear
MOLECULE TYPE: protein
US 08 764 870-11

Query Match	54, 38; Score 710; Lnk 4; Length 427;
Host Local Similarity	40, 38; Prod. No. 1, 70-66;
Matches	166; Conservative 136; Indels 48; Gaps 8

[illegible]

RESULT
 US 08 459, 489, 10
 Sequence 10, Application US/08459489
 Patent No. 5666574
 GENERAL INFORMATION:
 APPLICANT: David D. Moore et al.
 TITLE OF INVENTION: CAR RECEPTORS AND RELATED
 TITLE OF INVENTION: MOLECULARS AND METHODS
 NUMBER OF SEQUENCES: 10
 CROSS-REFERENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZITE: 02110 2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 50Z or 555X
 OPERATING SYSTEM: IBM P.C. DOS (Version 3.40)
 SOFTWARE: WordPerfect (Version 5.0)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08459, 489
 FILING DATE:
 CLASSIFICATION: 415
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/845, 450
 FILING DATE: February 26, 1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Paul T. Clark
 REGISTRATION NUMBER: 40,162
 REFERENCE/DOCKET NUMBER: 00780/126001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5000

TELEFAX: (617) 542-8900
TELEX: 200154
INFORMATION FOR SPO ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 448
TYPE: amino acid
STANDARDNESS: N/A
TOPOLOGY: linear
INS 08 459-489-10

Query Match	33.48;	Score	741;	DB	1;	Length	348;
Host Local Similarity	40.58;	Pred. No.	9, 90-65;				
Matches	157;	Conservative	59;	Mismatches	100;	Indels	72;
						Gaps	6

QY	41	CRVQDKTQVGVHVMTEGCKGFEPRAMKNMRLQVFKKGAETIKRLKROVQALRL	160
DB	11	GVVQVQATQVTHMALTEGCKGFEPRKRVKSKTCTPFGAGSEVYAKTQRRHVARRLQ	69
QY	101	KLTSGMKKEMIMSDVAEVEPRALIKRKRSHRTQPLQVGLTEDEYRMRLRLMAQMK	160
DB	70	KLTJAGMKKQMLLSAFLALIKRAQVABQAVTQV-----LSKQFEHLRTLLVAHTR	12
QY	161	TEPTTSHFKNFRLLPGVLISSGCELPESLQASREPRAMKSVQRKDLNLSVSLQKGEQ	228
DB	124	HMGIMFEQFVQFRRPAMLEHHQPLDPLAP-----	15
QY	221	SVNMKRPVAGSGKFRPSLSLPRHMAQNTVMYKGLISAKVLSYPRKQPLTEQSLKGA	280
DB	154	-----VLALVTHPMDINTEPMVGVKTRKQPLVRSLSLDEQSLKGA	108
QY	281	FEICQIRETNFVNAETQWEGGRISYCLTDIAG-GPQDILDEPMIAKPYMKKQLAPPE	349
DB	199	VEICIVNTFTFCQTNQNLGSLPRYLTEDSAHAGVQVEPFLDFFRRGTLRKQVQDEP	258
QY	340	YVLMOALSLSPDRPGVQGHVNVQDLOQFATTKSTIKSNRQVPAHREPLKIMAMITE	399
DB	259	YVLLAAMALFSPDRPGVQVQDEIDOLOEEMALLOSYSIKGVQRPRDPRFLYAKILGLAE	318
QY	400	FAT-----PIPMQF	408
DB	419	LRSTINEAVQVQVQHLQGLISAMMPLIQGL	446

RESULT 3
 US-08/458-686-10
 Sequence 10. Application US/08/458686
 Patent No. 5710017
 GENERAL INFORMATION:
 APPLICANT: David D. Moore et al.
 TITLE OF INVENTION: GAC RECEPTORS AND RELATED
 TITLE OF INVENTION: MOLECULES AND METHODS
 NUMBER OF SCIENTISTS: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FISH & Richardson
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 50Z or 55SX
 OPERATING SYSTEM: IBM P.C. DOS (Version 4.30)
 SOFTWARE: Microsoft (Version 5.0)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US-08/458,686
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/843,350
 FILING DATE: February 26, 1992
 ATTORNEY/AGENT INFORMATION:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC DOS/MS DOS
 SOFTWARE: Patient ID Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/737,746B
 FILING DATE: 19910740
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Schwartz, Carl R.
 REGISTRATION NUMBER: 29,447
 REFERENCE/DOCKET NUMBER: 96,296,2195,2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 414-277-5215
 TELEFAX: 414-277-5774
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 467 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 MOTIF TYPE: Prohibit
 HYPOHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: RAT
 PUBLICATION INFORMATION:
 AUTHOR: Burmester, James K.
 ADDRES: Macdon, No. 526019490
 AUTHORS: Macdon, No. 526019490
 TITLE: Isolation and expression of rat
 JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
 VOLUME: 85
 PAGES: 1005-1009
 DATE: February-1988
 US 07 737-746B-4

Query Match 27.0% Score 590; DR 1; Length 467;
 Post Local Similarity 46.4% Prod. No. 1.2e-50;
 Matches 186; Conservative 65; Mismatches 125; Indels 48; Gaps 8;
 74 KLRGPKKACETKTRKROGACLRKRLKESQKKRKMIMSDAVERRALIKRKRSEK 134
 1 KETTPK MACKTKKRNRRQACRLKRYVLCMKKEFTTDEHYGKKRMMKKEFEA 59
 134 GTGPTVQGLTERDQRMRLRLMDQMKTTTTSHTKRLGVLS--SSGFLPSLGAP 191
 60 LKSLDKR LSEPGQHLTALILAHKRYDPAADPRDPPVPMKSGTSYSP--P 114
 192 SREAAKMSQVRRKGLSKLV-----STQDGEKGSVMNKKYADSGKELPSLAP 242
 115 TLEPSNSSSSSSSLYFTSLDMEKSGSNLDNGESD--DVSYTLDSPLMDR 169
 244 KAMSTYMKKGLISPAKYSPKRLPIEDQSLKGAAPFLQDRENVFNATETVWEG 402
 170 EADVSTSGKAVIGAKKIPGRKLTSDQVLKSSALVIMKSSALVIMKSSALVIMKSS 229
 403 PLASYLEPT--AGGPGQALLERMKRYMKRLQDDEEYVLMQALISFSDRVRVGR 359
 240 SOLKRYIVTVSKAHELELEELKRYVLMKRLINLEHEVILMAKILVSDPRVQDA 269
 460 RYVIGLQDPAITKSYLHGNQPAHPLPFLKIMAGTEP----- 400
 290 KIVKALQDKSLNTQYTRKRPVPGSGIYAKKIQKALAKSLNPKNSKQYKSLSGVE 349
 401 -ATPKQGLRG 410
 430 NSMKLTPLVLEVGR 463
 RESULT 6

US-08-430-518-2
 Sequence 2; Application US/08330518
 Patent No. 5607967
 GENERAL INFORMATION:
 APPLICANT: Fittman, Brian
 APPLICANT: Hollaway, M. Katherine
 APPLICANT: Rodan, Gideon
 APPLICANT: Schmidt, Axel
 APPLICANT: Vogel, Robert
 TITLE OF INVENTION: USE OF RECEPTOR POTENTIATORS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESS: Merck & Co., Inc.
 STREET: 126 East Lincoln Avenue
 CITY: Rahway
 STATE: New Jersey
 COUNTRY: US
 Z11: 07065-0907
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC DOS/MS DOS
 SOFTWARE: Patient ID Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/430,518
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Dolan, Catherine A.
 REGISTRATION NUMBER: 46,502
 REFERENCE/DOCKET NUMBER: 19416
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-4283
 TELEFAX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 461 amino acids
 TYPE: amino acid
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 MOTIF TYPE: Prohibit
 HYPOHETICAL: NO
 ANTI-SENSE: NO
 US-08-430-518-2

Query Match 21.6% Score 472; DR 1; Length 461;
 Post Local Similarity 28.3% Prod. No. 9.2e-49;
 Matches 122; Conservative 80; Mismatches 119; Indels 110; Gaps 14;
 37 GPTQVQVCTKATQVHENVMTGEGKQPPKAMKRNALIKYER-KGAEFTKTRPGQ 95
 1 GELTQVQVCTKATQVHENVMTGEGKQPPKAMKRNALIKYER-KGAEFTKTRPGQ 142
 83 GELTQVQVCTKATQVHENVMTGEGKQPPKAMKRNALIKYER-KGAEFTKTRPGQ 142
 96 AGRKQKTESQKKRKMIMSDAVERRALIKRKRSEKTV--LQAVGQ----- 142
 143 GCHLRKRLKESQKKRKMIMSDAVERRALIKRKRSEKTV--LQAVGQ----- 142
 144 -----LTPEDQRMRLRLMDQMKTTTTSHTKRLGVLS--SSGFLPSLGAP 191
 201 SPQSTASQNSQSGPGQVLTAAQELMDQVLAQVQGNKSS-- 244
 184 LPSLAPKSLREAAKMSQVRRKGLSKLV-----STQDGEKGSVMNKKYADSGKELPSLAP 242
 245 -----DQPKVTWP-----LQAD--PGRDAKQGRFA--HF 271
 244 ADMSTYMKKGLISPAKYSPKRLPIEDQSLKGAAPFLQDRENVFNATETVWEG 402
 272 TELATISVQETVPAKQVNGPLQAREQVIMKSSALVIMKSSALVIMKSSALVIMKSS 229
 402 -CHLSYGLTD--TAGGPGQALLERMKRYMKRLQDDEEYVLMQALISFSDRVRVGR 359
 429 FLKDFYKSLQDPRAGLQVEFTNPTFESRMRRLGLDQVYALALAMSDAPNVQDE 488

QY 359 HRVVDQIOEFAITLKSYTECNKRPQAHF--LFLKIMAMLT-----EFA----- 401
DB 389 PGRVADLPVVEALLSTRIRKPDQDLRPPMLKLVSLRTLSVSHSEVVALRLQDKK 448
QY 402 -TPLMOELFGI 411
DB 449 LPPILSEIMDV 459

RESULT 7
US-08-330-283-2
Sequence 2, Application US/08330283
Patent No. 5679518
GENERAL INFORMATION:
APPLICANT: Friedman, Eitan
APPLICANT: Holloway, M. Katharine
APPLICANT: Rodan, Gideon
APPLICANT: Rutledge, Su Jane
APPLICANT: Schmidt, Azriel
APPLICANT: Vogel, Robert
TITLE OF INVENTION: METHOD FOR FINDING RECEPTOR POTENTIATORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 East Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330-283
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Dolan, Catherine A.
REGISTRATION NUMBER: 36,502
REFERENCE/DOCKET NUMBER: 19327
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4283
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-330-283-2

Query Match 21.6% Score 472: DB 1: Length 461:
Best Local Similarity 28.3% Pseq 39: Indels 110: Gaps 14:
Matches 122: Conservative 80: Mismatches 119:

QY 37 GPQICRVGCDKATGYHENVMTGCGKGFRRAMKRNALRCPEK-KGAGEITRKRQOQ 95
DB 83 GHIELCVGDKKASGFHYNYLSCGCGKGFRRSVRGARRACRGGTGCOMAFMRKQO 142
QY 96 AGLRKCLSSKKKMMIMDEAVERRALIKRKKSRGTQ---PLGVGQ----- 142
DB 143 QCKLRCKEAGMRQCVSLFEVIRKKK--TRQOQOESOSOSOPVGPQSSSSASGPGA 200
QY 143 -----LTFEORMIRLMDQKRTDTTFSHKRNRLPGVLSGCE 183
DB 201 SPQGSFAGSGSGSGGCVQVITAAQHLMIQQLVAALQCKNRKRSF----- 244

QY 184 LPESLQAPSRREFAKWSQVRKDIQSLKVSLOLRGDDGSVMNKRPPADSGCKEIRSLPHM 243
DB 245 -----DQKRVTPW-----LGAD-----PQSRHARQORFA---HF 271
QY 244 ADMSTYFKGJLISFAKVSYFRDLPIDQISLKGAEFLQULFNIVFNAETGTEWC-- 301
DB 272 TELATISVOELVDRAKQVPGFLQCRQDIALKASTIEIMLFTARKYNIEET---ECII 328
QY 302 --GRLSYGLD-TAGGPDQLLEPMLKRYMLKRIQHEPPYVLMQALSLSPHSPGVLC 358
DB 329 FLKDFYKSKDDFHRAGLOVEFIINLPEFSRPMRRIGLIDARVALLIINIHSALRPVVOE 388
QY 359 HRVVDQIOEFAITLKSYTECNKRPQAHF--LFLKIMAMLT-----EFA----- 401
DB 389 PGRVADLPVVEALLSTRIRKPDQDLRPPMLKLVSLRTLSVSHSEVVALRLQDKK 448
QY 402 -TPLMOELFGI 411
DB 449 LPPILSEIMDV 459

RESULT 8
US-08-646-248-2
Sequence 2, Application US/08646248
Patent No. 5939422
GENERAL INFORMATION:
APPLICANT: Friedman, Eitan
APPLICANT: Holloway, M. Katharine
APPLICANT: Rodan, Gideon
APPLICANT: Rutledge, Su Jane
APPLICANT: Schmidt, Azriel
APPLICANT: Vogel, Robert
TITLE OF INVENTION: METHOD FOR FINDING RECEPTOR POTENTIATORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 East Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,248
FILING DATE: 14-MAY-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/330,283
FILING DATE: 27-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Dolan, Catherine A.
REGISTRATION NUMBER: 36,502
REFERENCE/DOCKET NUMBER: 19327
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4283
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-646-248-2

FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Quagliato, Carol S.
REGISTRATION NUMBER: 35,330
REFERENCE/DOCKET NUMBER: 19116 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3809
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 461 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-13931-2

Query Match 21.6%; Score 472; DB 5; Length 461;
Best Local Similarity 28.3%; Pred. No. 9,28-39;
Matches 122; Conservative 80; Mismatches 119; Indels 110; Gaps 14;

QY 37 GQIOTRVGDKATGTHFNVMTCGCKGFFRRAMKRNALRCFPR-KGACETITRKTROCO 95
DB 83 GHEILRVGCDKSGPHYVNLDCGCKGFFRRSVRGARVACRGCGTQMDATMRKCO 142
QY 96 ACRLKRCLESGKKKEMINSDAEVERRALIKRKKSERGTG---PLGVG----- 142
DB 143 QCRLRKCKAGKRECVLSEEDIRKK--IRKQOOGOSOSOSIVGPGSSSSASGPGA 200
QY 143 -----ITFEORPMIRELMDAQKTFDTTFSHKRNRLPGVLSGCE 183
DB 201 SPGSEASGSGSGEGVQLTAQELMIQULVAQLOGCKRHS- 244
QY 184 LPESLQAPREFAAKWSOVRKDCSLKVSLOLKGEDGSVMNKKPPADSGCKETISLLPHM 243
DB 245 -----DQPVTPW-----LGAD-----PSRDARQOQRA--HF 271
QY 244 ADMSTYMRKGIISFAKVISYFRDLPIEDQISLKGAFELCOLRNTFNATGTWEC-- 301
DB 272 TELALISVEIYDPAKQVGFQLOREDQIALKASTIEIMLETARKYNHET---ECIT 328
QY 302 --GRUSYCLD--TAGGFQOLLLEPMLEKHYMLKKIQLHEEYVLMQALISPSPPVNIQ 358
DB 329 FCKDFYSKDDPHRAGLOVEFIPFIFFSRAMRRLGLDDAEVALLIAINIFSADRPVQE 388
QY 359 HRVVDQLOQPAITLKSYIECNRPQPAHRF--LFLKIMAMLT-----EFA----- 401
DB 389 PGREVALQDPYVEALLSYTRKRPQDQLRPPEMLKIVLSLRTLSVSHSVOYVALRLQDK 448
QY 402 -TPLMQELFGI 411
DB 449 LPPLLSETMDV 459

RESULT 11
US-08-342-411A-2
Sequence 2, Application US/08342411A
Patent No. 5639616
GENERAL INFORMATION:
APPLICANT: LIAO, Shutsung
APPLICANT: SONG, Ching
TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR:
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX

COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,411A
FILING DATE: 18-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KITCHELL, BARBARA S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: ARCD154
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-342-411A-2

Query Match 21.5%; Score 470; DB 1; Length 460;
Best Local Similarity 28.3%; Pred. No. 1,46-38;
Matches 122; Conservative 80; Mismatches 119; Indels 110; Gaps 14;

QY 37 GQIOTRVGDKATGTHFNVMTCGCKGFFRRAMKRNALRCFPR-KGACETITRKTROCO 95
DB 82 GHEILRVGCDKSGPHYVNLDCGCKGFFRRSVRGARVACRGCGTQMDATMRKCO 141
QY 96 ACRLKRCLESGKKKEMINSDAEVERRALIKRKKSERGTG---PLGVG----- 142
DB 142 QCRLRKCKAGKRECVLSEEDIRKK--IRKQOOGOSOSOSIVGPGSSSSASGPGA 199
QY 143 -----ITFEORPMIRELMDAQKTFDTTFSHKRNRLPGVLSGCE 183
DB 200 SPGSEASGSGSGEGVQLTAQELMIQULVAQLOGCKRHS- 243
QY 184 LPESLQAPREFAAKWSOVRKDCSLKVSLOLKGEDGSVMNKKPPADSGCKETISLLPHM 243
DB 244 -----DQPVTPW-----LGAD-----PSRDARQOQRA--HF 270
QY 244 ADMSTYMRKGIISFAKVISYFRDLPIEDQISLKGAFELCOLRNTFNATGTWEC-- 301
DB 271 TELALISVEIYDPAKQVGFQLOREDQIALKASTIEIMLETARKYNHET---ECIT 327
QY 302 --GRUSYCLD--TAGGFQOLLLEPMLEKHYMLKKIQLHEEYVLMQALISPSPPVNIQ 358
DB 328 FCKDFYSKDDPHRAGLOVEFIPFIFFSRAMRRLGLDDAEVALLIAINIFSADRPVQE 387
QY 359 HRVVDQLOQPAITLKSYIECNRPQPAHRF--LFLKIMAMLT-----EFA----- 401
DB 389 PGREVALQDPYVEALLSYTRKRPQDQLRPPEMLKIVLSLRTLSVSHSVOYVALRLQDK 447
QY 402 -TPLMQELFGI 411
DB 448 LPPLLSETMDV 458

RESULT 12
US-08-372-652-3
Sequence 3, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Wonja
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING

Guilty Match	21.2%	Score	464	104.2%	Length	446
Post-Local Similarity	28.5%	Prod. No.	5,564	48		
Matrices 121	Conservative	78	Mismatches	122	Indels	104
					Gaps	14

[illegible]

Only Match	21.28	Score	464	104.58	Length	446	
Best Local Similarity	28.58	Pred. No.	5	50-38			
Matches	121	Conservative	78	Mismatches	122		
				Indels	104	Gaps	14

[illegible]

10b	128	HSRCHTPILOTYMRKCKOCHETLHKKRBOACHBHEVLESEUJLHKK	1KKOUEHEGHAHVS	184
07	142	KTGTUPLACVUOLITETGUMMMELMDIAOMKPTUTTFNSHKFKRLGVLS	SSGCLLPESL	188
10b	185	LPHKSSSPQULHPILOSPLOMPEKLVAAOCCNRSES		223
07	189	QAPSEEEAAKKSVOVKRDLSTKVSJLDRDEMSWYNKRPALASGKRELSL	PHKMAWST	248
10b	224	DRKJTPHMPAP	DHSHRBAKQORRA	HTEHIAL 255
07	245	YMRKPLISPAKVISTYPRHPLTHOJLSLKKADAPLAVLRNYEN	AEFTGIMKGRHSY	306
10b	256	VSVOELVDFAPALACFQOLSHEDJIALHKSALFVMMLETSKRYNNGSSTTP	ELKPFST	314
07	307	GLEPFA GCFQULLLEPKHFMKTKLQJHEEYVYVMOALSIESPBRQVLO	HRVVOI	365
10b	315	NHEDFAKAGLOVERINTLFFSRKMMLOJONAFALIALSIFSAKRIVQULOV	ERK	374
07	366	GQFATLTKSVTEGNQVPAHPLFKIMONTFEAT		405
10b	375	QHTVEALIAVVSITHBPL	DRMFPEMLKIVLSIKLSVSHSEGVAPALDOKKLPPL	432
07	406	QELPPL 411		
10b	433	SEIWDV 438		

Search completed: JULY 11, 2001, 09:10:45
Job time: 51 sec


```

34  LuValGlyGlyProGlnIleCysArgValCysGlyAspLysAlaThrGly 50
|||||
170 AACCTGGAGCTCCCAATCTGCGCTGATGTGGGACCAAGCGCCACTGGC 219
51  TyrHisPheAsnValMetThrCysGlyGlyCysLysGlyPhePheArgTrp 67
|||||
220 TATCACTTCATATGATGACATGTGACAGATGCAAGCGCTTTTCAGGAG 269
67  GAlaMetLysArgAsnAlaArgLeuArgCysProPheArgLysGlyAlaC 84
|||||
270 GCCATGAAACGCCACGCCCGCTCAGGTGCGCTTCGGAAGCGCGCT 319
84  ysgLulIethrArgLysThrArgArgGlnCysGlnAlaCysArgLeuArg 100
|||||
320 GCGAGATCACCCCGAAGACCGCGGACAGTCCGACGCTGCCCGCTGGCG 369
101  LysCysLeuGlnSerGlyMetLysLysGlnMetIleMetSerAspGlnAl 117
|||||
370 AAGTGGCTGGACAGCGCGCATGAAGAGAGATATGATGCCACGAGCG 419
117  aValGluArgArgAlaLeuIleLysArgLysLysSerGluArgTrpC 134
|||||
420 CGTGGAGAGACGCGCGGCTTGTATCAAGCGGAGAAAGTGAACGAGAG 469
134  LyrThrGlnProLeuGlyValGlnGlyLeuThrGlnGlnArgMetMet 150
|||||
470 GGACTGACCGCCTGGAGAGTGCAGGGGCTGACAGAGAGCGGAGTATG 519
151  IleArgGluLeuMetAspAlaGlnMetLysThrPheAspThrThrPhe 167
|||||
520 ATAGGAGAGTGTATGAGCGCTCAGATGAAGAACTTTGACCTACCTTCT 569
167  rHisPheLysAsnPheArgLeuProGlyValLeuSerSerGlyCysGlu 184
|||||
570 CCATTTCAAAGATTTCCGCTGCCAGCGGCTTACGACGTGGCGAGT 619
184  euProGlnSerLeuGlnAlaProSerArgGlnGlnAlaAlaLysTrpSer 200
|||||
620 TGGCAGAGTCTCTGCAGGCCCATCGAGGAGAGAGTGCAGAGTGGAGC 669
201  GlnValArgLysAspLeuCysSerLeuLysValSerLeuGlnLeuArg 217
|||||
670 CAGGTCCGGAAGATCTGCTCTTGAAGTCTCTCTGAGCGTGGG 719
217  yGluAspGlySerValIlePasnTrpLysProProAlaAspSerGlyGly 234
|||||
720 GGAGGATGGCAGTGTGGAACACCAACCCCGAGCGAGTGGCGGGA 769
234  ysgLulIethrSerLeuLeuProHisMetAlaAspMetSerThrTrpMet 280
|||||
770 AAGAGATCTTCTCCCTGCTCCCGACATGGCTGACATGTCACCTTAC 819
251  PheLysGlyIleIleSerPheAlaLysValIleSerTrpPheArgAsp 267
|||||
820 TTCAAGGCGATCATCAGCTTTGCCAAGTATCTCTACTTCAGGAGCT 869
267  uProIleGlnAspGlnIleSerLeuLeuLysGlyAlaAlaPheGlnLeu 284
|||||
870 GCCCATGAGACACAGATCTCCCTGCTGAGAGGCGGCTTTCACCTGT 919
284  ysgLulLeuArgPheAsnThrValPheAsnAlaGlnTrpGlyTrpGln 300
|||||
920 GTAACATGCAATTCACACAGTGTTCACACCGGAGACTGGAACCTGGAG 969
301  CysGlyArgLeuSerTrpCysLeuGluAspThrAlaGlyGlyPheGln 317
|||||
970 TGTGGCGCGGTGTCTACTGCTTGGAGAGTACTGACAGTGGCTTCAGCA 1019
317  nLeuLeuLeuGlnProMetLeuLysPheHisTrpMetLeuLysLeuG 334
|||||
1020 ACTTCTACTGAGCGCATGCTGAAGATTCACATGCTGACAGAGTGGC 1069

```

```

334  InLeuHisGlnGlnLysIleValIleLeuMetGlnAlaIleSerLeuPheSer 350
|||||
1070 ACCCTGATGAGGAGAGATGTGCTGATGACAGCCCATCTCCCTTCTGC 1119
351  ProAspArgProGlyValLeuGlnHisArgValAlaAsnGlnLeuGln 367
|||||
1120 CCAGACGCGCGTGGGTGTGTCGACGACCGCGGTGTGTCGACGACGAC 1169
367  uGlnPheAlaIleThrLeuLysSerTrpIleGlnCysAsnArgProGln 384
|||||
1170 GCAATTCGCGATTCAGTCAAGTCTGACATTCATGATGACATCGCCAGC 1219
384  rAlaHisArgPheLeuPheLeuLysIleMetAlaMetLeuThrGln 399
|||||
1220 CTGCTCATAGGTCTTGTCTGTAAGATCATGCTATGCTACCGAGCTC 1269
399  .....
1270 CCGACATTCATGCTGTCGACACCGCGGCTGCTGCGCATTCAGAGCAT 1319
400  .....PheAlaThrProLeuMetGlnLysLeuPheGlyIleThrGlyS 414
|||||
1320 ACACCCCTTGTCTACGCGCCCTCATGACAGAGTGTGTCGATCAGAGTA 1369
414  er 414
||
1370 GC 1371
seq_name: qb_pat2:E32702
seq_documentation_block: 1422 bp  DNA PAT 07-FEB-2001
LOCUS E32702
DEFINITION Novel nuclear receptor protein, its gene and utilization function.
ACCESSION E32702
VERSION E32702.1 GI:13026809
KEYWORDS JP 1999127872-A/3.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1422)
AUTHORS Jun, Y.-Y., S.-S. and Naito.
TITLE Novel nuclear receptor protein, its gene and utilization function
JOURNAL JAPAN TOBACCO INC
PATENT JP 1999127872-A 3 18-MAY-1999;
COMMENT OS Homo sapiens (human)
PN JP 1999127872-A/3
PD 18-MAY-1999
PF 07-AUG-1998 JP 1998224172
PR
PI JUN YAMAMOTO, YUTAKA SAITO, TAKAYUKI NAITO
PC C12N15/09, C07K14/72, C07K16/28, C12N1/21, C12P21/02, C12O1/68, PC
G01N33/53,
PG G01N33/566/(C12N15/09, C12R1:91), (C12N1/21, C12R1:19), (C12P21/02, C12R1:19),
PC C12N15/00, (C12N15/00, C12R1:91)
CC
FH Key location/Qualifiers
FT CDS 1..1422
FEATURES
source location/Qualifiers
1..1422
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 336 a 395 c 414 g 277 t
ORIGIN
alignment_scores:
Quality: 2167.00 Length: 434
Ratio: 5.234 Gaps: 1
Percent Similarity: 95.392 Percent Identity: 95.392
alignment_block:

```


Align seq 1/1 to: H32702 from: 1 to: 1422

seq_name	seq	seq_length	seq_type	seq_desc	seq_date
1418	CGAATTCGATTTACGACATGTTTCAACCGGCAACATGAACTGGAC	1017	cds	1418	1999
401	CysGlyAlaGlyLeuSerTyrCysLeuGluAspThrAlaGlyGlyProGlnGlu	417	cds	401	1999
1018	TCGGGCGAGCTGCTGCTACTGCTTGGAAAGACACTGACGGTGGCTTCACAA	1067	cds	1018	1999
417	MetLeuGluGluGluProMetLeuLeuGlyPheHisSerMetLeuGlySerMet	444	cds	417	1999
1098	ACTTTTACGAGAGCGGATGCTGCAAAATTCGATACATGCTGCGAAGAAATCTG	1117	cds	1098	1999
434	IndLeuHisGluGluGluGlyValValLeuMetGlnAlaAlaSerLeuPheSer	450	cds	434	1999
1118	ACTTTTACGAGAGCGGATGCTGCAAAATTCGATACATGCTGCGAAGAAATCTG	1167	cds	1118	1999
451	ProAspArgProGluValLeuGluHisGlyValValAspGluGluGluGlu	467	cds	451	1999
1168	CGAAGACGCTCAGGCTGCTGACAGACGCGCTGCTGGACACAGCTGGAGAA	1217	cds	1168	1999
467	MetLeuAlaAlaLeuLeuLeuGlySerTyrLeuLeuGlyAsnAlaProGluLeu	484	cds	467	1999
1218	CGAATTCGATTTACGACATGTTTCAACCGGCAACATGAACTGGAC	1267	cds	1218	1999
484	LeuAlaHisArgPheLeuGluGlyLeuMetAlaAlaLeuThrGlu...	499	cds	484	1999
1298	CGCTCATAGGTTTGTGTTTTCGAAATCATGCTATGCTGCAACAGCTC	1317	cds	1298	1999
499	499	cds	499	1999
1318	CGAATTCGATTTACGACATGTTTCAACCGGCAACATGAACTGGAC	1367	cds	1318	1999
409PheAlaThrProGluMetGlnGluLeuPheGlyLeuThrGly	414	cds	409	1999
1468	ACGCGCGCTTCTGACGCGCTCATGACGAGAGTGTGGAGATACACAGTA	1417	cds	1468	1999
414	or 414		cds	414	1999
1418	GC 1419		cds	1418	1999


```

/db_xref="GI:351138"
/translation="MEVPRKESMNADEHVCEDTESYPRKPSYNADPEVGGPOTGVY
GKATGCHFNVMTCDEGCKFFRRAMKRNALRCPFRGACETIRKTRRQCAELKRC
LESQMKEMTSDAEVERALLIKRKSERTGPOLVQDLTEORRMILRLMDAOKM
TFDTFSHFNRLPLGVLSGCELPESLQAPSRKASQVRDLCSLVSOLRBE
DGSVMNKRPPADSGKEIFSLPLMADMTYMERGISLPAKVISYFDLPIEDQISLL
KGAFFELCOLRENTVENAETGECRLSYCLEETAGFPQDLEPMLKRYHMLKRLK
LHEEYVLMQALISFSPDPGVLOHRYVDOLEQFATLLKSYIECNRPQAHNRFLELK
IMMLTFLKRNAGTORRLRIODIHFFATPLMDLELFGITGS"

BASE COUNT      509 a      593 c      605 g      439 t
ORIGIN

alignment_scores:
    Quality: 2167.00      Length: 434
    Ratio: 5.234      Gaps: 1
    Percent Similarity: 95.392      Percent Identity: 95.392

alignment_block:
US-09-276-935b-14 x AF061056 ..

Align seg 1/1 to: AF061056 from: 1 to: 2146

1  LeuGlValArgProLysSerTyrPasnHisAlaaspPheValHisCy 17
|||||
304 CTGGAGGTGAGAGCCCAAGAAAGCTGAAACCATGCTGACTTTGTACACTG 353

17  sgluasPthrGluSerValProGlyLysProSerValAsnAlaaspGlu 34
|||||
354 TGAGGACACAGAGTCTGCTCGTGAAGAACCCCATGTCACAGCGAGATGAG 403

34  LuValAlGlyProGlnInIleCysArgValCysGlyAspLysAlaThrGly 50
|||||
404 AAGTCGAGGTCCCAAAATCTGCGTGTATGTGGGACAAAGCGCACTGGC 453

51  TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPhePheArg 67
|||||
454 TATCACTTCATATGTCATGACATGTGAAGATGCAAGGCTTTTCAGAG 503

67  GAlaMetLysArgAsnAlaArgLeuArgCysProPheArgLysGlyAla 84
|||||
504 GGCATGAAACGCAACGCCGCTGAGGCTCCCTTCGGAAGAGGCGCT 553

84  ysgLueLethrArgLysThrArgArgLncysIleAlaCysArgLueuArg 100
|||||
554 GCGAGATCAACCCGCAAGACCCGCGACAGTCCGAGGCTGCCCTGGCG 603

101  LysCysLeuGluSerGlyMetLysLysGluMetIleMetSerAspGlu 117
|||||
604 AAGTCGCTGAGAGCGCGCATGAAAGAGATATCATGTCGACAGAGC 653

117  aValArgLueArgArgAlaLeuIleLysArgLysLysSerGluArgThr 134
|||||
654 CCGGAGAGAGAGCGCGCTTGTATCAAGCGAAGAAAGTGAAGCGACAG 703

134  LyrThrGlnProLeuGlyValGlnGlyLeuThrGluGluIleArgMet 150
|||||
704 GGACTCAGGCACACTGGGAGTGCAGGGCTGACACAGAGAGCGAGATGAT 753

151  IleArgGluLeuMetAspAlaGluMetLysThrPheAspThrThrPhe 167
|||||
754 ATCAGGAGAGTGTATGACCTTCAGATGAAACCTTGACACTGACTTCT 803

167  rHisPheLysAsnPheArgLueuProGlyValLeuSerSerGlyCysGlu 184
|||||
804 CCATTTCAAGAAATTTCCGCTGCCAGGGCTGCTTACAGATGGCTGAG 853

184  eueProLysSerLeuGlnAlaProSerArgGluGluAlaAlaLysThrP 200
|||||
854 TGCCAGAGTCTTGCAGAGCCCATCAGGAGAAACCTGCAAGTGGAGC 903

201  GluValArgLysAspLysSerLeuLysValSerLeuGlnLeuArgGly 217
|||||

```

```

904 CAGTCGCGAAGATCTGTGCTTTGAAGGTCTCTCTGACGATGGCGGG 953
217 ygluasPcySerValTyrPasnTyrLysProProAlaaspSerGlyL 234
|||||
954 GGAGATGGCAGTGTCTGAACTACAAACCCCAAGCCAGTGGCGGCA 1003

234 ysgLueLethrArgLysThrArgArgLncysIleAlaCysArgLue 250
|||||
1004 AAGAGATCTTCTGCTGCTGCCCAATGGCTGACATGCTGACATCACT 1053

251 PheLysGlyIleIleSerPheAlaLysValIleSerTyrPheArgP 267
|||||
1054 TTCAGAGGATCATCATGCTTTGCCAAAGTCACTCTGCTACTTCAAG 1103

267 uProLleGluaspGlnIleSerLeuLeuLysGlyAlaAlaPheVal 284
|||||
1104 GCCATCGAGAGACAGATCTCTGCTGAGAGGCGCGCTTTCAGAGT 1153

284 ysgLueLethrPheAsnThrValPheAsnAlaGluThrGlyThrP 300
|||||
1154 GTCAACTGAGATTCACACACTGTTCAACGCGGAGACTGGAACTGG 1203

301 CysGlyArgLysSerTyrCysLeuGluaspThrAlaGlyPheGln 317
|||||
1204 TGTGGCGCGCTGTCTGCTGCTGAGAGACACTGACGTTGGCTTCAG 1253

317 nLeuLeuLeuGluProMetLeuLysPheHisTyrMetLeuLysLeu 334
|||||
1254 ACTTCTACTGAGCCCATGCTGAAATTCCTACATCTCTGAAAGACT 1303

334 nLeuHisIleGluGlnIleTyrValIleuMetGlnAlaIleSerLe 350
|||||
1304 AGCTGCATAGAGAGAGATGTGTGCTGATGACAGGCGCATCTCTTC 1353

351 ProAspArgProGlyValIleuGlnHisArgValAlaaspGlnLe 367
|||||
1354 CCAGACCGCGCGAGTGTGCTGACAGACCGCGTGGAGACACTGCA 1403

367 uGlnPheAlaIleThrLeuLysSerTyrIleGluCysAsnArgPro 384
|||||
1404 GCATTCGCGCATTACTGTAAGTCTACATTCATTAATGCAATGCG 1453

384 rAlaHisIleArgPheLeuPheLeuLysIleMetAlaMetLeuThr 399
|||||
1454 CTGCTCATAGGTTCTTGTCTGAAAGATCATGCTATGCTACGAGCT 1503

399 ..... 399
1504 CCGAGATCAATGCTGACACACCGCGGCTGCTGCGATTCAGAGACT 1553

400 .....PheAlaThrProLeuMetGlnIleuLeuPheGlyIleThr 414
|||||
1554 ACACCCCTTCTGACCGCCCTCATGACAGAGTGTGGCATCAGAGTA 1603

414 et 414
||
1604 GC 1605

seq_name: gb_pr4:AF084644

seq_documentation_block:
LOCUS      AF084644      2802 bp      mRNA
DEFINITION Homo sapiens orphan nuclear receptor (P4K2) mRNA, complete cds.
ACCESSION  AF084644
VERSION    AF084644.1 GI:3769536
KEYWORDS
SOURCE
ORGANISM  human;
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 2802)
            Bertilsson,G., Heidrich,J., Svensson,K., Asmou,M., Jendberg,
            Sydow-Backman,M., Ohlsson,R., Postlund,H., Blomquist,P. and

```


400PheAlaThrProLeuMetGlnGluLeuPheGlyIleThrGlyS 414
 1427 ACACCCCTTGGTACCCCTCATGACGAGCTTGTTCGGCATCACAGTA 1476
 414 er 414
 1477 GC 1478
 seq_name: gb_pt4:AF084645
 seq_documentation_block:
 LOCUS AF084645 2905 bp mRNA PRI 20-OCT-1998
 DEFINITION Homo sapiens orphan nuclear receptor (PARI) mRNA, complete cds.
 ACCESSION AF084645
 VERSION AF084645.1 GI:3769538
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS Bertilsson, G., Heidrich, J., Svensson, K., Asman, M., Jendeberg, L.,
 Sydow-Backman, M., Ohlsson, R., Postlind, H., Blomquist, P., and
 Berkenstam, A.
 TITLE Identification of a human nuclear receptor defines a new signaling
 pathway for CYP3A induction
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (21), 12208-12213 (1998)
 MEDLINE 98445350
 REFERENCE 2 (bases 1 to 2905)
 AUTHORS Bertilsson, G., Asman, M., Blomquist, P., and Berkenstam, A.
 TITLE Direct Submission
 JOURNAL Submitted (19-AUG-1998) Cell and Molecular Biology, Medical Nobel
 Institute, Karolinska Institute, Doktorringen 2, Stockholm 17177,
 Sweden
 FEATURES
 source location/Qualifiers
 1..2905
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="liver"
 1..2905
 /gene="PARI"
 280..1584
 /gene="PARI"
 /codon_start=1
 /product="Orphan nuclear receptor"
 /protein_id="AAC64558.1"
 /db_xref="GI:3769539"
 /translation="MEVRPKESMNHADPVHCEPESVPGKPSVNDPEVGGPOTCRVC
 GKATGTHPNVATCGCGKGFPRKAKRNKALRCPRKACETTRTRGCGACRIKRC
 LESGKKRKLMSDEAVERRALIRKKSERTGTPLGVGGLTEEDRKMTRELMDQMK
 FEDTFSHKKNRPLGVLSSGCELPESLQAPSREBAWMSVVRKDLCKLSLOLGE
 DGSVWYKPPADSGGKEIFSLPHMADSTYFKGISFAKVISYFRPLIEDQISLQ
 KGAFELCOLRNTVENAETGMECGRLSYCELDAGSFOOGLLEPMLKPHYMKLKK
 LHEEYVLMQATSLSPSPGVLQHRVAVDLOEOPALITKSYTECNRQPAHRELFELK
 IMAMTLELRISNAGTORLKRQDIHPRATPLMOELPRTGS"
 BASE COUNT 765 a 727 c 778 g 635 t
 ORIGIN
 alignment_scores:
 quality: 2167.00 length: 434
 ratio: 5.234 gaps: 1
 percent similarity: 95.392 percent identity: 95.392
 alignment_block:
 US-09-276-935b-14 x AF084645
 Align seq 1/1 to: AF084645 from: 1 to: 2905
 1 LeuGluValArgProIySerTyrPAsnHisAlaAspPheValIhscy 17
 280 CTGAGCGTGGACCCCAAGCTGACCATCTTGTACACTG 329

17 scIuAspThrGluSerValProGlyIySerSerValAsnAlaAspGlu 44
 330 TAGAGCAACAAAGTCTGTCTGGAAAGCCAGTGTCAAGCGCAGTARG 479
 34 LuValGlyIyProIyLeuCyAsArgValCyGlyIyAspIyAlaThrGly 50
 380 AAGTCGAGGTCGCCCAATCTGCCGTATGTGGGACAAAGCCACTGGC 429
 51 TyrHisPheAsnValMetThrCysGluGlyCysIyGlyPhePheArgAr 67
 430 TATCACTTCATATCATGATCATGATGATGATGATGATGATGATGATG 479
 67 gAlaMetIyArgAsnAlaArgLeuArgGlyProPheArgIyValArg 84
 480 GGCATGAAACCAACCCGCGGTGAGGTGCGCTTCCGGAAGGACCC 529
 84 YscIuIleThrArgIyThrArgArgIyIncYsGluAlaIyAsArgLeuArg 100
 530 GCGAGATCAACCGGAGAGACCCGCGCAGCTGCCAGGCTCGCGCTCGC 579
 101 IyScYsLeuGluSerGlyMetIyIySyluMetIleMetSerAspGluAl 117
 580 AAGTCCTGACAGCGCGCATCAAGAGAGATGATGATGATGATGATGATG 629
 117 AValGluGluArgArgAlaLeuIleIyArgIyIySyluSerArgIyVal 134
 630 CTGGAGAGAGAGCGCGCTTATCATCAAGCGGAACAAATGCAACGAC 679
 134 YThrGlnProLeuGluIyValGluGlyLeuThrGluGluIyArgMetMet 150
 680 GGAATCAACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
 151 IleArgGluLeuMetAspAlaGluMetIyStrPheAspThrThrPhe 167
 730 ATCAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
 167 rHisPheIySAsnPheArgLeuProGlyIyValLeuSerSerGlyIyGlu 184
 780 CATTTCAGAAATTTCCGCTGCGCAGAGGCTGTACCAATGAGTGGAG 829
 184 euProGluSerLeuGluAlaProSerArgIyGluAlaIyAlaIyStrPhe 200
 830 TGGCAGAGTCTTGAGGCCCATCGAGGAGAGAGAGAGAGAGAGAGAG 879
 201 GlnValArgIyAspLeuGlySerLeuIyValSerLeuGlnIleuArg 217
 880 CAGTCGCGGAAATCTGTCTTGAAGGTCTGTGCAGCTGCGGAG 929
 217 YGluAspGlySerValITrPAsnIyIySerProAlaAspSerGlyIy 234
 930 GGGGAGTGGAGCTGTGAACTACAAACCCGAGCGAGAGAGAGAGAG 979
 234 YscIuIlePheSerLeuLeuProIyMetAlaAspMetSerThrTyrMet 250
 980 AAGAGATCTTCTCCGCTGCGCAGATGATGATGATGATGATGATGATG 1029
 251 PheIyGlyIleIleSerPheAlaIyValIleSerTyrPheArgAspLe 267
 1030 TTCAAGGCAATCATCACTTCCCAAGATCATCTCTACTTGAAGGACT 1079
 267 uProIleGluAspGlnIleSerLeuLeuIyGlyIyAlaIyPheGluArg 284
 1080 GCCCATGAGAGACCAATCTCCCTGCTGAAGGCGCGCTTTCAGACT 1129
 284 YscIuLeuArgPheAsnThrValPheAsnAlaGluThrGlyThrTrpGlu 406
 1130 GTCACTGAGATTCACACAGCTTTCACGCGGAGAGTGAACCTGGAG 1179
 301 CysGlyArgLeuSerTyrCysIuGluAspThrAlaGlyIyGlyIyGlu 417
 1180 TGTGGCGGCTGTCTACTGCTTGAAGAGACACTGCGAGTCTTCAGGA 1229


```

|||||
1075 GCCCATCGAGGAGACATCTCCCTGCTGAGGCGCGCTTGAGACTGT 1124
284 yslInleuArqPheAsnThrValPheAsnAlaGluThrGlyThrTrpGlu 300
|||||
1125 GTCACTGACATTCACACAGTGTTCACGCGGACATGCAACTGGCAG 1174
301 CysGlyArqLeuSerTyrCysLeuGluAspThrAlaGlyGlyPheGln 317
|||||
1175 TGTGGCGGCTGTCTACTGCTTGGAGACACATGCGAGGTGCTTCAGCA 1224
317 leuLeuLeuGluProMetLeuLysPheHisTyrMetLeuLysLysLeu 334
|||||
1225 ACTCTACTGAGAGCCATGCTGAATTCACATCACTGCAAAAGCTGC 1274
334 leuLeuHisGluGluGluTyrValLeuMetGlnAlaLeuSerLeuPheSer 350
|||||
1275 AGCTGATGAGGAGAGATATGCTGATGACAGCCATCTCCTCTTCTTC 1324
351 ProAspArqProGlyValLeuGlnHisArqValValAspGlnLeuGlu 367
|||||
1325 CCAGAGCGCCGACAGTGTCTGACACCGCGTGTGGACAGCTGCAGGA 1374
367 uGlnPheAlaIleThrLeuLysSerTyrIleGluCysAsnArqProGlnP 384
|||||
1375 GCATTCGCCATTACTCTGAGTCTACATTCATTCATTCGCCCCAGC 1424
384 roAlaHisArqPheLeuPheLeuLysIleMetAlaMetLeuThrGlu... 399
|||||
1425 CTCTCATAGTCTTGTCTGATGATCATGCTATGCTCACCGAGCTC 1474
399 ..... 499
1475 CGCAGCATTAATGCTCAGCACACCGCGCTGCTCGCATTCAGAGCAT 1524
400 ..... PheAlaThrProLeuMetGlnGluLeuPheGlyIleThrGly 414
|||||
1525 ACACGCCCTTGTACGCCCTCATGACAGGATGTGTTCGCATCAGGTA 1574
414 er 414
|||||
1575 GC 1576
seq_name: qb_pat2:E32708
seq_documentation_block:
LOCUS E32708 3243 bp DNA PAT 07-FEB-2001
DEFINITION Novel nuclear receptor protein, its gene and utilization thereof.
ACCESSION E32708
VERSION E32708.1 GI:13026815
KEYWORDS JP 1999127872-A/9.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 3243)
JUN, T.Y.S.S. and Naito.
Novel nuclear receptor protein, its gene and utilization thereof
Patent: JP 1999127872-A 9-18-MAY-1999;
JAPAN TOBACCO INC
OS Homo sapiens (human)
PN JP 1999127872-A/9
PD 18-MAY-1999
PF 07-AUG-1998 JP 1998224172
PR
P1 JUN YAMAMOTO, YUTAKA SAITO, TAKAYUKI NAITO
PC C12N15/09, C07K14/72, C07K16/28, C12N1/21, C12P21/02, C12Q1/68, PC
G01N33/53,
PC G01N33/566//C12N15/09, C12R1:91), (C12N1/21, C12R1:19), PC
(C12P21/02, C12R1:19),
PC C12N15/00, (C12N15/00, C12R1:91)
CC
FH Key Location/Qualifiers

```

```

FEATURES             FT            CDS           (344)..(1765).
     source           1..3243
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT      838 a      806 c      862 g      737 t
ORIGIN
alignment_scores:
    quality: 2167.00      Length: 434
    Ratio: 5.234      Gaps: 1
    Percent Similarity: 95.392      Percent Identity: 95.392
alignment_block:
us-09-276-935b-14 x E32708
Align seq 1/1 to: E32708 from: 1 to: 3243
1 leuGluValArqProLysGluSerTrpAsnHisAlaAspPheValHisCys 17
|||||
461 CTGGAGGTGACACCCAAAGAAAGCTGGAACCATGCTGACTTGTACACTG 510
17 sGluAspThrGluSerValProGlyLysProSerValAsnAlaAspGluG 34
|||||
511 TAGGACATACAGTCTCTTCTCTGGAAGCCCACTGTCAACGACATAGAG 560
34 luValGlyGlyProGlnIleCysArqValCysGlyAspLysAlaThrGly 50
|||||
561 AAGTGGAGAGTCCCAAAATCTCCGCTGATATGGGACAAAGCCACTGAC 610
51 TyrHisPheAsnValIleMetThrCysGluGlyCysLysGlyPhePheArqAr 67
|||||
611 TATCACTTCAATGTCATGATGATGATGATGATGATGATGATGATGATG 660
67 gAlaMetLysArqAsnAlaArqLeuArqCysProPheArqLysGlyAlaC 84
|||||
661 GGCATGAAATGCAACGCCCGGCTGAGTGGCCCTTCCGGAAGGCGGCT 710
84 yslInuIleThrArqLysThrArqGlnCysGlnAlaCysArqLeuArq 100
|||||
711 GCGAGATCACCCGGAAGACCGCGGACAGTGCACAGGCTGTGCGCTGCG 760
101 LysCysLeuGlnSerGlyMetLysLysGlnMetIleMetSerAspGlnAla 117
|||||
761 AAGTCCCTGAGAGCGGCATGAAGAGGATGATCATGTCGACAGAGC 810
117 aValGluGluArqArqAlaLeuIleLysArqLysLysSerGluArqThrG 134
|||||
811 CGTGGAGGACAGCGGCGCTTATCATCAGCGGAAGAAAGTCAACCGACAC 860
134 lyThrGlnProLeuGluValGlnGlyIleThrGluGluGlnArqMetAla 150
|||||
861 GSACTTACCCACATGGGAGTGCAGGGGCTGACAGAGACAGAGGATGATG 910
151 leuArqGluLeuMetAspAlaGlnMetLysThrPheAspThrThrPheSe 167
|||||
911 ATCAGAGACATGATGACCTCAGTGAAGAAACCTTGGACACACACTCTTC 960
167 rHisPheLysAsnPheArqLeuProGlyValLeuAsrSerGlyCysGlu 184
|||||
961 CCATTTCAGAAATTTCCGGCTTCGCAGAGGCTGCTTACCACTGCTGCAGAT 1010
184 euProGluSerLeuGlnAlaProSerArqGluGlnAlaAlaLysIlePheSer 200
|||||
1011 TGGCAGAGTCTGTGAGCCCATGAGGAGGAAGAAACCTGCAAGTGGAGC 1060
201 GlnValArqLysAspLeuGlySerLeuLysValSerLeuGlnLeuArqG 217
|||||
1061 CAGTCCGGAAGATCTGTCTCTTTGAAGAGTCTCTCTCACTCGCGGG 1110
217 ysluAspGlySerValIlePheAsnTyrLysProProAlaAspSerGlyGly 234
|||||

```


alignment block:

US-09-276-935b-14 x HSAJ9936

Align seg 1/1 to: HSAJ9936 from: 1 to: 4448

```

1 LeuGluValArgProLysGluSerThrPheAsnAlaGlnThrGlyThrProGlu 17
1840 CTGGAGGTGAGACCCCAAGCAAGCTGCAACCATGACTGCTTGTACACCTC 1889
17 scLysAspThrGluSerValProGlyLysProSerValAsnAlaAspGlnG 34
1890 TGAGGACACAGAGTCTGCTGCTGGAAGCCGACGTGCAACGCGAGATGAG 1939
34 LuValGlyGlyProGlnIleCysArgValCysGlyAspLysAlaThrGly 50
1940 AAGTCGAGGTCCCAATCTGCGGTGATGTGGGACCAAGGCCACTGGC 1989
51 TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPhePheArg 67
1990 TATCACTTCATGTCATGACATGTGAAGATGCAAGCCCTTTTTCAGCAG 2039
67 gAlaMetLysArgAsnAlaArgLeuArgCysProPheArgLysGlyAla 84
2040 GGCATGAAACGCAACGCCCGCTGAGTGCCCTTCGGAAGCGCGCT 2089
84 YscGluIleThrArgLysThrArgArgGlnCysGlnAlaCysAlaGluArg 100
2090 GTCAGATCACCCGGAAGACCCGCGACAGTGCAGCGCTGCGCCCTGCGC 2139
101 LysCysLeuGluSerGlyMetLysLysGluMetIleMetSerAspGlu 117
2140 AAGTGGCTGGAGAGCGGCATGAAGATGATCATGTCGACAGCAGGC 2189
117 aValGluGluArgArgAlaIleuIleLysArgLysLysSerGluArgThr 134
2190 CTTGGAGACAGAGCGCGCTTGTGATCAAGCGAGAAAGTGAACGAGCAG 2239
134 LysThrGlnProLeuGlyValGlnGlyLeuThrGlnGlnIleArgMetMet 150
2240 GGCATCAGCCAGTGGAGTGCAGGCGCTGACAGAGGACGCGGATGATG 2289
151 IleArgGluLeuMetAspAlaGlnMetLysThrPheAspThrPhe 167
2290 ATCAGGAGCTGATGAGCGCTGATGAATGAATCATGTCGACAGCAGGC 2339
167 rHisPheLysAsnPheArgLeuProGlyValLeuSerSerGlyCysGlu 184
2340 CCAATTCAAGCAATTCGCGCTGCGCAGCGGCTTAGCAGTGGCTCGAGT 2389
184 euProGluSerLeuGlnAlaProSerArgGlnGluAlaAlaLysTyrSer 200
2390 TGGCAGAGTCTGCGAGGCCCATCGAGGAAAGAGTGCACACTGAGC 2439
201 GlnValArgLysAspLeuCysSerLeuLysValSerLeuGlnLeuArg 217
2440 CAGGTGCGGAAGATGCTGCTCTTTCAAGTCTCTCTGCACTGCGCGG 2489
217 yGluAspGlySerValTrpAsnTrpLysProProAlaAspSerGly 234
2490 GAGAGTGGAGTGTGTGGAACACCAACCCGACGCGCAGTGGCGGGA 2539
234 YscGluIlePheSerLeuLeuProHisMetAlaAspMetSerThrMet 250
2540 AAGAGATCTTCTGCTGCTGCGCCACATGCTGACATGCAACCTACATG 2589
251 PheLysGlyIleIleSerPheAlaLysValIleSerThrPheArgAsp 267
2590 TTCAGAGCATCATCAGCTTGCAGAAAGTATCTCTCTACTTCAGAGCAT 2639
267 uProIleGluAspGlnIleSerLeuLeuLysGlyAlaAlaPheGluLeu 284
2640 GCCTATGAGAGCCAGATCTCCTGCTGGAAGGCGCGCTTTCGAGGTGT 2689

```

```

284 YscGlnLeuArgPheAsnThrValPheAsnAlaGlnThrGlyThrProGlu 300
2690 GTCAACTGAGATTCAACACAGATGTTCAACGCGAGAGCTGGAACCTGGAG 2739
301 CysGlyAlaIlePheSerTyrCysLeuGluAspThrAlaGlyGlyPheGln 317
2740 TGTGGCGGGCTGTCTACTGCTTGGTCAAGACACTGCACGTGGATTCAGTA 2789
317 nLeuLeuLeuGluProMetLeuLysPheHisTyrMetLeuLysLysLeuG 334
2790 ACTTCTACTGAGAGCCCATCTGTAATTCACATGATGCTGGAAGAACTGC 2839
334 nLeuHisGlnGluGlyTyrValLeuMetGlnAlaIleSerLeuPheSer 350
2840 AGCTGATGAGAGGAGATGTGCTGATGCAAGCCATCTCCCTCTTTCG 2889
351 ProAspArgProGlyValLeuGlnHisArgValValAspGlnLeuGln 367
2890 CCAGACGCCCGCAGGTGTCTGTCAGACAGCCGCTGCTGACACCTTGAGGA 2939
367 uGlnPheAlaIleThrLeuLysSerTyrIleGluCysAsnArgProGln 384
2940 GCAATTCGCGATTACTGTAAGTCTACATGTAATGCAATCGCGCCAGC 2989
384 rAlaHisArgPheLeuPheLeuLysIleMetAlaMetLeuThrGlu 399
2990 CTGCTCATGCTTGTGTTCTGTAACATCATGCTATGCTACCGAGCTC 3039
399 ..... 399
3040 CGCAGATCAATCTCAGCACACCAGCGGCTGCTGCGCATCGACAGAT 3089
400 .....PheAlaThrProLeuMetGlnGluLeuPheGlyIleThrGlyS 414
3090 ACACCCCTTGTGACCCCTCATGAGAGAGTGTGGCGCATCAGAGTA 3139
414 er 414
3140 GC 3141
seq_name: qb_p9: HSAJ9937
seq documentation block:
LOCUS HSAJ9937 4337 bp mRNA PRI 09-SEP-1999
DEFINITION Homo sapiens mRNA for nuclear hormone receptor PRR2.
ACCESSION AJ009937
VERSION AJ009937.1 GI:5852066
KEYWORDS nuclear hormone receptor; orphan nuclear receptor; PRR2 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 4337)
REFERENCE
1 Heard, D.J., Holloway, J., Hansen, C., Tommerup, N., Aagaard, L. and
Vissing, H.
Identification of a novel protein isoform of the human nuclear
hormone receptor PXR/SXR and localization to chromosome 3p14.1 -
13.3
Eur. J. Hum. Genet. In press
JOURNAL
2 (bases 1 to 4337)
AUTHORS Heard, D.J.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-1998) Heard D.J., Novo Nordisk A/S, Molecular
Genetics, Novo Allé, DK-2880, Bagsvaerd, DENMARK
FEATURES
location/Qualifiers
1..4337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="q12-q13.3"
/tissue_type="liver"
/issue_type="intestine"

```


/Issue . 433- "color"
 1771 . . 4033
 /gene "PRR2"
 1771 . . 4033
 /gene "PRR2"
 1771 . . 4033
 /note "Isomorph with human A/B domain. Translation
 initiates at cdb-1771 5' fold less efficiently than at
 cdb-1840"
 /odon_start 1
 /transl_except (pos:1771 . . 1773,aa:Met)
 /product "mature hormone receptor PRR2-c"
 /protein_id "AF55493.1"
 /db_xref "GI:5852068"
 /translation "MPHRYGKRNIPNSRGREPNLEVGRKSNMTHDVGDEPTE
 SYGKISVADDEEVGIVGTCGVCGRKATYHNPVETETGTFRRKRRKARLRTE
 RGVGTEHRTTRGQGVLRKRTPESEKNNIKNSDEVERKALIKRKKSKYRTPDGL
 LKQVETEDGRNRIEELMDADKRTDTTESHKRVSLDGRGDSWNRKPRKPS
 GCEIPESTLDHMAIMSYMKGIISPAKVIISYRQLIDQDLSIKGAAPFELQRPENT
 VNAETGHWETGRSLYCEITAGGQVLLLEPMKPYMKLQDHEEYVLMQVLSI
 PSNDGRVQDINVDLQDQFATLKSTLEPNNRQVAHRLPLKIMATLETLSINQ
 HFGRLKIDYHPEATPLMDLEPLTGS"
 1840 . . 4033

[illegible]

```

/feature "truncated receptor"
/feature_start 1
/product "antibiotic resistance receptor PRB2_H"
/product_id "CAB55494.1"
/db_xref "ref: 585,2065"
/translation "MTEKAKGFFBRAMKNAKICLPKGGACDILKKTRKQVYAKI
RCKLESTMKNNMIMSDAVERRKALIKKSKSRGIPDLYVQGLLESDRIMILRIMDA
QKLEIDTELSHKRQKRSVSLKCEGSSWNNKPRVDSNGKRLFSLIDMLMSTYMK
GLISPAKYSTYRLLPDEDDSLKGAATFLGQLRNTVNAHGTWEGHLSYELH
TACGQVQDILFPMIRHYMKLQIDLEEVYMQATSLFSDRGIVLQHRVVDLQVQ
PAATKSYLTCNRVPIHRELFLKIMATLELRSINAHIVQRLRLQIDIPPAVPLMQ
ELFGITTS"
```

[illegible][illegible]

seq_name	db_om	AF182217
seq_documentation_block:		
LOCUS	AF182217	1895 bp mRNA MAM 19-APR-2000
DEFINITION	Oryctolagus cuniculus pregnancy x receptor (NR112) mRNA, complete cds.	
ACCESSION	AF182217	
VERSION	AF182217.1	GI:5853353
KEYWORDS		
SOURCE		
ORGANISM		
	rabbit.	
	Oryctolagus cuniculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.	
REFERENCE	1 (bases 1 to 1895)	
AUTHORS	Savas, U., Wester, M. R., Griffin, K. J., and Johnson, E. F.	
TITLE	Rabbit pregnancy x receptor is activated by ritampicin	
JOURNAL	Drug Metab. Dispos. 28 (5), 529-537 (2000)	
MEDLINE	20235257	
REFERENCE	2 (bases 1 to 1895)	
AUTHORS	Savas, U., Wester, M. R., Griffin, K. J., and Johnson, E. F.	
TITLE	Direct submission	
JOURNAL	Submitted (01-SEP-1999) Division of Biochemistry, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA	
location/Qualifiers		
FEATURES		
source	1..1895	
	/organism="Oryctolagus cuniculus"	
	/strain="bred New Zealand white rabbit"	
	/db_xref="taxon:9986"	
	/tissue_type="kidney"	
	1..1895	
	/gene="NR112"	
	/note="pxr"	
	257..1492	
	/gene="NR112"	
	/note="pxr: nuclear receptor subfamily 1 group 1 member 2; zinc-finger protein; transcription factor"	
	/codon_start=1	
	/product="pregnane x receptor"	
	/protein_id="AAD54426.1"	
	/db_xref="gi:5853354"	
	/translation="MGQSPITSADEFESPTQRYGQDKANGYHENVLTGECQKGFPR	
	TVKNNARISQCPKRGACEITKTRKROCACALRKCIESGKKEMISQVAAORAL	
	KRRKRNHQAQPPMOGLIGQRMILIELMLAQKIPDTTSHRNFLPVLSGCGE	
	IPSELQATEELGAKRMHQLQELSLTKSLDRLREDSYVMYTPPAPDSCKLSLLE	
	HLADSTYMERGILINFAKIVSTIRFLPLFLDILSLKATLTLCILRENTVNAETGIV	
	ECGLASTVERPEEGCPDLDPLIKRHYMKQIDKEEYVLMQALSLFSPDRGVV	

ORFVDDLOEEFAFLTKAY IECSSPPQTHRLFIKIMAVLTFLRTIMAGHITQRIILRLR
DTHPEALPLMRELFSTDD**

BASE COUNT 434 a 556 c 535 g 370 t

ORIGIN

alignment_scores:

	Quality: 1732-50	Length: 432
Percent Similarity: 89.352	Ratio: 4.488	Gaps: 2
Percent Identity: 75.926		

alignment_block:

US-09-276-935H-14 x AF182217 ..

Align seg 1/1 to: AF182217 from: 1 to: 1895

```
1 LeuGIuValArqProLysGluSerTPAsnHisAlaAspPheValHisGly 17
|||||
191 CTGAGAGTCTGATATCTGAGGCACTGGGAAACAGACCTGGCTTCGGCGTG 240
17 SGUASPTnHgluSerValProGluLysProSerValAsnAlaAspGluG 34
|||||
241 TCATGAGACGACCTCCATGGGTGTAAGAACCCACCATATGAGTAAATGAG 290
|||||
34 LuvalGlyGlyProGluHisLeuSerValLysGlyAspLysAlaThrGly 50
|||||
291 AG...GAGGGCCCGACACTTGGCGTGTATCCGGGACAGAGCCCAATGCT 337
|||||
51 TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPheArqAl 67
|||||
338 TACCACCTCAAGCTCCTGACATGTGAAGCTGCAAGGCTTTTCAGAG 387
|||||
67 GAlAMeLysArqAsnAlaArqLeuArqCysProPheArqLysGlyAla 84
|||||
388 GACCGTGAACCGCAACCCCGGCTGAGATGCCCTTCGGAGAGGGCCCT 437
|||||
84 ysgluIethrArqLysThrArqArgGlnCysGlnAlaCysArqLeuArq 100
|||||
438 GCGAGATCACTCGGAAACGACCGGCGGCTGAGCGCTCTCGCTGCGG 487
|||||
101 LysCysLeuGluSerGlyMetLysLysGluMetLleMetSerAspGluAl 117
|||||
488 AAGTGGCTGGAGAGCGGCATGAAAGAGGAATATCATCTGCGCATGTGCT 537
|||||
117 aValGluGluArqArqAlaLeuLleLysArqLysSerGluArqThrG 134
|||||
538 CGTGGCGACAGAGAGCGGCTCTGATCAAGAGAGAGAAACAGAAACGATG 587
|||||
134 LyrThrGlnProLeuGlyValGlnGlyLeuThrGluGluGlnArqMet 150
|||||
588 AAGCTCGGCGCGCGGATGCAAGGGCTGAGCGAGGACGAGCGATGATA 637
|||||
151 IleArgGluLeuMetAspAlaGlnMetLysThrPheAspThrThrPheS 167
|||||
638 ATGAGAGGCTGATGACGCGCTCAAGATGAAACCTTTGATCAGCACTTTC 687
|||||
167 rHisPheLysAsnPheArqLeuProGluValLleuSerSecGlyCysGlu 184
|||||
688 CCAATTCAACAATTTTGGCTGCGCGAGAGTGGCTTGGAGTGGCTGTGAGA 737
|||||
184 euProGluSerLeuGlnAlaProSerArqGlnGluAlaAlaLysTyrSer 200
|||||
738 TTCCACATCTCTGACGGCTTAAGGAGAGAGAGAGGCTGTGACATGATA 787
|||||
201 GluValArqLysAspLeuCysSerLeuLysValSerLeuGluLeuArqAl 217
|||||
217 ygluAspGlySerValTyrAsnTyrLysProProAlaAspSerGlyGly 234
|||||
838 GGAAGAGCGGAGCTCTGGAACATAACGCCGCCAGCTGACAGACAGATGGGA 887
|||||
234 ysgluIethrSerLeuLeuProHisMetAlaAspMetLysThrTyrMet 250
```



```

454 CCGAGAGCTGCTGGCAGCTGGTGTGAGATTCAGAAATCTCTCAGACCTT 503
191 oSeratggtguguaiaaialaLysTrpseroiNaValarqlysaSpluLeuYys 208
504 AACGGAGAGAGAGAGCTGGCAGCTGAGAGACGATCCAGAGAGAGAGCTGG 553
208 eLeuLysValSerLeuGlnLeuArqgLYGlaspeLySerValTrpAsn 224
554 CCATGAGAGCTCTCTCTGACAGCTGGCGGAGAGAGAGAGAGCTTGGAGC 603
225 TyTrsProAlaAspSerLyLysGlyLysGlyLeuPheSerLeuLeuP 241
604 TATAGGCCCCCAGCTGACAGAGAGTGGGAAAAAGCTTTCTCTCTGCTGCC 653
241 oHisMetAlaAspMetSerThrTyrMetLysGlyLysGlyLeuPhe 258
654 CCAGTGGCTGACATGTCAACCTACATCTTCAAGGCATCATCACTTTG 703
258 LaLysValLysSerTyrPheArqAspLeuProLeuLysPheLeuSer 274
704 GTAAAGTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 753
275 LeuLeuLysGlyAlaAlaPheGlnLeuLeuYysGlnLeuArqPheAsnTrp 291
754 TTGCTGAAGAGAGAGAGAGCTGGCAGCTGCTCTCTCTCTCTCTCTCTCT 803
291 LpHeaSnalagLthrGlyThrTrpGluYysGlyArqLeuSerTyrCysL 308
804 GTTCAACGAGAGAGAGAGAGCTGGCAGCTGCTCTCTCTCTCTCTCTCTCT 853
308 euGluAspThrAlaGlyGlyPheGlnGlnLeuLeuLeuLeuProMetLeu 324
854 TCGAAACACCTCAAGAGAGAGCTTCCAGCAACTCTCTGGACCCCTGGCG 903
325 LysPheHisTyrMetLeuLysLysLeuGlnLeuHisGlnGlnLysVal 341
904 AATTTCAGTACATGCTGAAGAGAGCTGAGAGAGAGAGAGAGAGAGAGCT 953
341 lleuMetGlnAlaLysSerLeuPheSerProAspArqProGlyValLeuG 358
954 CGTGAAGAGAGAGAGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1003
358 LnhIsArqValValaAspGlnLeuGlnGlnGlnPheAlaLysThrLeuYys 374
1004 AGCGTGAAGTGGAGAGAGCTTCCAGAGAGAGCTTCCAGAGAGAGAGAG 1053
375 SerTyrLysGlnCysAsnArqProGlnProAlaHisArqPheLeuPhe 391
1054 GCTTCATGATGAGTGGAGAGAGCTGGCGGAGAGAGAGAGAGAGAGAGCT 1103
391 uLysIleMetAlaMetLeuThrGln.....PheAlaThrProLeu 404
1104 GAAGATCATGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1153
405 MetGlnGlnLeuPheGlyLysThr 412
1204 ATCGGAGAGCTGTTCAGCAGCAGCA 1227
seq_name: qb_r01:AF031814
seq_documentation_block:
LOCUS AF031814 1709 bp mRNA RDB 07-FEB-1998
DEFINITION Mus musculus pregnancy X receptor mRNA, complete cds.
ACCESSION AF031814
VERSION AF031814.1 GI:2852328
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 1709)
Jones, S.A., McKee, J.T., Wade, L., Staudinger, J.L., Watson, M.A.,
Zetterstrom, R.H., Perlmann, T., and Lehmann, J.M.
An orphan nuclear receptor activated by pregnanes defines a novel
steroid signaling pathway
Cell 92 (1), 73-82 (1998)
JOURNAL
MEDLINE
9814945
REFERENCE
AUTHORS
2 (bases 1 to 1709)
Jones, S.A., McKee, J.T., Wade, L., Staudinger, J.L., Watson, M.A.,
Zetterstrom, R.H., Perlmann, T., and Lehmann, J.M.
Direct Submission
Submitted (29-OCT-1997) Department of Molecular Endocrinology,
Glaxo Wellcome, 5 Moore Drive, RTP, NC 27709, USA
FEATURES
SOURCE
1. 1709
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="liver"
151..1446
/function="steroid signaling"
/note="pregnane-activated nuclear receptor"
/todon_start=1
/product="pregnane X receptor"
/db_xref="GI:2852328"
/translation="MRPESMSHVGIVOCHEADSALEPPINVEHEGGLCTIKVTCGDK
ANGYHFNMTCEGCGFFRRAMRNKLPKPRKCTGCTTRTRGQAIRKIKLES
GMKKEHMSDAVEORRALIKRRKKRITADPPGAGGTEEDVALIDELMIVOMUTFD
TTSHEKRLRPAVFHSGCELPEDLASLLEDAVMSQIMKIRVIMISIGRIGES
IMVQPPSKSDGKEIIPILPHLADVSTYMKGVINAKVSYRQDPIEHOJSLKGA
TEPCGLRENTMDTEGTWECGRLAVCEDEMGFQKLLIDPLMKFHEMKLOLHK
EFVIVLQALSLFSDRPGVYORSVVOLODERVALTKVAIEGSRVFAHPPIETKIMA
VLEIHSINAOQTOOLLRIDSHSPVATPIPLMOELFFSTING"
BASE COUNT 449 a 444 c 468 g 348 t
ORIGIN
alignment_scores:
Quality: 1684.50 Length: 412
Ratio: 4.468 Gaps: 2
Percent Similarity: 87.269 Percent Identity: 73.380
alignment_block:
US-09-276-935B-14 X AF031814
Align seq 1/1 to: AF031814 from: 1 to: 1709
1 LeuGluValArgProLysLeuSerTrpAsnHisAlaAspPheValHisCys 17
145 CTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 194
17 sGluAspThrGlnSerValProGlyLysProSerValAlaSnalAspGlnG 34
195 TGAAGAGAGAGAGAGCTTGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 241
34 LuValGlyLysProGlnLysAspArqValCysGlyAspLysAlaThrCys 50
242 AAGATGAGAGCTTTTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 291
51 TyrHisPheAsnValMetThrCysGlnGlnCysLysGlyPhePheArgAr 67
292 TACCACTTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 341
67 gAlaMetLysArqAsnAlaArgLeuArqCysProPheArqLysGlyAlaC 84
342 GGCCATGAAGAGAGAGAGCTTGGCTGAGGAGAGAGAGAGAGAGAGAGAGCT 391
84 YsgLysGlyThrArgGlnCysGlnAlaCysArqLeuPhe 100

```



```

|||||.....|||||.....|||||.....|||||.....|||||.....|||
1209 ACCTGGTCGAGGACGACTGCCTGTATGACAGCCCATCACCCTCTTCTC 1258

351 PRAAFAPRPRGLVYALenGINHISArqVALIASPILNLeuchI 367
|||||.....|||||.....|||||.....|||||.....|||||.....|||
1259 CCAGATCGCCCTGCCCTGTTCACTACCTAGCTGGTAGACACGACTGCAGGA 1308

367 ucInpheAlalIEtherLysSerTyrlleGIucYSASNArProGlnP 384
|||||.....|||||.....|||||.....|||||.....|||||.....|||
1309 GGATTTCCTCACCTTGACAGCCATCATCCAGTAAGTACTACGCCCTAATC 1358

384 rGalHisrrPhelEulephelusIIemElAMelleuthroU... 399
|||||.....|||||.....|||||.....|||||.....|||||.....|||
1359 CTGCACACAGGTTCTGTCTCTGAAGATCATGCTGTCTCTCCACCCAGCTG 1408

399 ..... 399

1409 CGCAGTATATATGCCAGACAGCCACGAGCTACTGCGATCCAGACAC 1458

400 .....PheAlatrProLeuMetGlInglutaphGlyllethr 412
|||||.....|||||.....|||||.....|||||.....|||||.....|||
1459 GCACCCCTTGGCACACCTTCATCATGAGGAGTTATTCACACACAGC 1504

seq_name= qb_ov:XLRNAONST
seq_documentation_block:
LOCUS XLRNAONST 1638 bp mRNA VRT 02-FEB-1994
DEFINITION X.laevis mRNA for orphan nuclear receptor of steroid/thyroid superfamily.
ACCESSION X75163
VERSION X75163.1 GI:410517
KEYWORDS steroid/thyroid hormone receptor superfamily.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE Smith,D.P., Mason,C.S., Jones,E.A. and Old,R.W. A novel nuclear receptor superfamily member in Xenopus that associates with RXR and shares extensive sequence similarity to the mammalian vitamin D3 receptor Nucleic Acids Res. 22 (1), 66-71 (1994)
JOURNAL MEDLINE 94173664
AUTHORS 2 (bases 1 to 1638)
TITLE Smith.D. Direct Submission
FEATURES Source Submitted (22-SEP-1993) D. Smith, University of Warwick, Dept of Biological Sciences, Gibbet Hill Road, Coventry CV4 7AL, UK
source Location/Qualifiers
1..1638 /organism="Xenopus laevis"
/db_xref="taxon:8355"
/drv_stage="gastrula (stage 1)"
/cclone_lib="cDNA library of gastrula (stage 1)"
245..1405 /codon_start=1
/product="orphan nuclear receptor of steroid/thyroid superfamily"
/protein_id="CAA53006.1"
/db_xref="gi:410518"
/translating="MMKVQETLVLEEEFEEDASNGITGDENKDKPKCRAGCGDAATGYHNMAATECGEGFRFRAWKRNLTPPPNSGVINKSRNRKQMGELKKCLDIDMKRLMSDADEQRALKRKHKTKLPPTPGNSLTPEQHFTGVNAVTRTFPDNNFKLNRPPIRRSDPTDPQSSEATKLPHSIDVTWIKGISAKPLFKSLDI EDIALKLSVAESVIRPNVINSDTNWVGEPFYDEMVLGF KQLFLPLELYRIHMKMLNLOSEEYAMMALSIASFARIPDCVMERKTOKIHRHAIATLKPIIOSRPFSPGNRIITYPKIMLCITFLTIVNDIHQSKLLIIDVIDPATLMREVEGSE"
```


Thu Jul 12 14:23:33 2001

us-09-276-935b-14.p2n.rge

Page 19

•
•
•

7


```

151 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 200
67 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 240
201 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 280
84 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 320
251 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 360
101 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 400
401 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 440
117 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 480
451 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 520
134 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 560
401 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 600
151 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 640
451 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 680
167 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 720
501 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 760
184 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 800
951 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 840
201 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 880
601 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 920
217 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 960
951 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1000
434 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1040
701 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1080
251 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1120
751 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1160
267 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1200
801 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1240
284 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1280
451 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1320
401 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1360
901 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1400
417 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1440
951 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1480
434 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1520
1001 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1560
451 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1600
1051 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 1640

```

```

467 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 184
1101 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 220
484 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 256
1151 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 292
499 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 328
1201 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 364
400 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 400
1251 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 436
414 TATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 472
1301 GATATCTTAAATTCATGCAATCTGCAAGCAATGCAATGCTTTCTTCAAGAC 508
seq_name: /SIBS/4/seq/geneseq/NA1999.LAT.AAX59967
seq_documentation_block:
ID AAX59967 standard: DNA: 1374 bp.
AC AAX59967:
AT 04 AUG-1999 (first entry)
DE DNA encoding an intranuclear receptor protein.
KW Human: intranuclear receptor protein; drug development; diagnosis;
   treatment; SS.
OS Homo sapiens.
FN JP11127872-A.
PD 18 MAY-1999.
PP 07 AUG-1998: 98JP-0224172.
PR 11-AUG-1997: 97JP-0230335.
PA (NISHI) JAPAN TOBACCO INC.
DE W01: 1999-350340/30.
DR P-PSDB: AAY15932.
XX
XX
XX New intranuclear receptor protein - useful for drug development and
   diagnosis and treatment of disease
XX
XX Claim 4: Page 24-25, 48pp; Japanese.
XX
XX The present sequence encodes a human intranuclear receptor protein.
   The nucleic acid sequence was isolated from a human adult cDNA
   library using a swedish AN23 derived probe. The protein can
   be used for the development of drugs and diagnosis and treatment
   of various diseases.
XX
XX Sequence 1374 bp: 629 A; 475 C; 406 G; 264 T; 0 other.
alignment_stats:
Quality: 2167.09 Length: 414
Ratio: 5.284 Gaps: 1
Percent Similarity: 95.392 Percent Identity: 95.392
alignment_block:
US-09-276-935b-14 x AAX59967
Align seq 1/1 to: AAX59967 from: 1 to: 1374

```


27-MAR-1998: 98US-0079593.
 XX (GLAXO) GLAXO GROUP LTD.
 XX Kilewer SA, Willson TM.
 XX WPI: 1999-601202/51.
 DR P-PSDB: AAY42691.
 XX
 PT New human pregnane X receptor, used to identify specific modulators and
 agents that induce expression of cytochrome P-450 mono-oxygenase
 XX
 PS Claim 4: Fig 1A-D; 6pp: English.
 XX
 CC The invention provides an isolated human nuclear receptor (designated
 CC pregnane X receptor, PXR) that binds to a cytochrome P-450 mono-oxygenase
 CC (CYP) promoter. The hPXR is used to identify its specific modulators,
 CC and compounds that induce CYP3A4 expression (i.e. to identify drug
 CC interactions, since CYP3A4 is involved in many biotransformations of
 CC drugs). The modulators are potentially useful for associating particular
 CC diseases and conditions with PXR and for treating such conditions.
 CC Antibodies raised against hPXR can be used for determination and
 CC purification of hPXR. The present sequence represents a DNA encoding the
 CC hPXR.
 SO Sequence 2146 BP: 509 A: 593 C: 605 G: 439 T: 0 other:

alignment_scores:
 Quality: 2167.00 Length: 434
 Ratio: 5.234 Gaps: 1
 Percent Similarity: 95.392 Percent Identity: 95.392

alignment_block:
 US-09-276-935b-14 x AA207997

Align seg 1/1 to: AA207997 from: 1 to: 2146

1 LeuGIuValArqProLysGluSerTTPAsnHISAlaAspPheValHisCy 17
 304 CTGGAGGTAGAGCCCAAGCAAGACGTGAAACATGCTTCTTACACCTG 353
 17 sguAspThrGluSerValProGlyLysProSerValAsnAlaAspGlu 34
 354 TGAAGACACAGAGCTCTGCTCTGAAAGCCCAATGTCACAGCACAATGAG 403
 34 luValGlyLysProGlnIleCysArqValCysGlyAspIysAlaThrGly 50
 404 AACCTGGAGGTCCCAATCTGCGTGTATGTGGGACCAAGCCCTGAGC 453
 51 TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPhePheArqAr 67
 454 TATCATTCAATGTGATGACATGTGAAGATGCAAGGCTTTTTCAGAG 503
 67 gAlaMetLysArqAsnAlaArqLeuArqCysProPheArqLysGlyAlaC 84
 504 GGCATGAAAGCAAGCCGCGCTGAGGTGCTCCCTTCGGAAGGGCGCTC 553
 84 ysgLulIethArqLysThrArqArqGlnCysGlnAlaCysArqLeuArq 100
 554 GCGAGATACACCGGAGACCGCGGACAGTGGCAGGCTCCGCTGGCG 603
 101 LysCysLeuGluSerGlyMetLysLysGluMetLleMetSerAspLul 117
 604 AAGTGGCTGGAGAGCGGCAATGAAGAGACATGATATGCCACGAGGC 653
 117 aValGluGluArqArqAlaLeuIleLysArqLysSerGluArqThrG 134
 654 GGTGGAGAGAGCGCGCTTGTGATCAAGCGGAAGAAAGTGAACGGACAG 703
 134 lYThrGlnProLeuGlyValGlnGlyLeuThrGlnGluGlnArqMetMet 150
 704 GGAATCGGACCTGGAGGTCCAGGGGCTGACAGAGAGGAGCGGATGATG 753

151 lIeArqGluLeuMetAspAlaGlnMetLysThrPheAspThrThrPheSe 167
 754 ATCAGAGACTGATGACGCTGACATGAAACCTTTGACACTACTCTTC 803
 167 rHisPheLysAsnPheArqLeuProGlyValLeuSerSerLysGlyLul 184
 804 CCATTTCAGAAATTTCCGCTGCCAGGGGTGCTTAGCAGTGGCTGGAGT 853
 184 euProGlnSerLeuGlnAlaProSerArqGluGluAlaAlaLysTrpSer 200
 854 TCCAGAGCTCTGACAGGCCCATCGAAGCAAGACCTGCCAAGTGGAGC 903
 201 GlnValArqLysAspLeuGlySerLeuLysValSerLeuGlnLeuArqG 217
 904 CAGGTCCGCAAGATCTGTGCTTTGAAGGCTCTCTGCAAGCTCCGCGG 953
 217 yGluAspGlySerValTTPAsnTrpLysProAlaAspSerGlyGlyL 234
 954 GGAGATGGCACTGTGTGAACTACAAACCCCGACCCAGCTGGTGGGA 1003
 234 ysgLulIethSerLeuLeuProHisMetAlaAspMetSerThrTyrMet 250
 1004 AAGAGATCTTCTCCCTGCTGCGCCACATGGCTGACATGCAACCTACATG 1053
 251 PheLysGlyIleIleSerPheAlaLysValIleSerTyrPheArqAspLe 267
 1054 TTCAAAGGCATATCATGCTTGGCAAGCATCTCTACTTACGGGACTT 1103
 267 uProIleGluAspGlnIleSerLeuLeuLysGlyValAlaPheGluLeu 284
 1104 GCCCATGACAGACCAAGATCTCCCTGCTGAAGGGGCGCTTTCAGAGCT 1153
 284 ysgLulIethArqPheAsnThrValIlePheAsnAlaGluThrGlyThrTrpGlu 300
 1154 GTCAACTGACATTCAAACACAGCTGTCAACGGCGAGACGTGAACCTGGAG 1203
 301 CysGlyArqLeuSerTyrCysLeuGluAspThrAlaGlyGlyPheGlnG 317
 1204 TGTGGCGGCTGCTCTACTGCTTGGAAACACTGGAGGGGCTTCCACGA 1253
 317 nLeuLeuLeuGluProMetLeuLysPheHisTyrMetLeuLysLysG 334
 1254 ACTTCTACTGAGGCCATGCTGAATTCACATGCTGGAAGAAAGCTGTC 1303
 334 lIleuHisGluGluGluTyrValIleuMetGlnAlaIleSerLeuPheSer 350
 1304 AGCTGCATGAGAGAGATATGTGTGATGACAGGCCATCTCCCTTCTTC 1353
 351 ProAspArqProGlyValLeuGlnHisArqValAlaAspGlnLeuGln 367
 1354 CCAGACCCGCGAGGAGTGTGTGACAGACCGCGTGGAGACCACTGCPAGA 1403
 367 uGlnPheAlaIleThrLeuLysSerTyrIleGluLysAsnArqProGluP 384
 1404 GCATATGGCATTTACTGTAAGTCTACATTCATGCAATGCAATGCGCCGAC 1453
 384 roAlaHisArqPheLeuPheLeuLysIleMetAlaMetLeuThrGlu... 399
 1454 GTCGTCAATGAGTTCTTGTCTTCTGAAGATGATGCTATGTCACGAGCTG 1503
 399 399
 1504 CCGAGCATCAATCTCAGACACCCAGCGGCTGCGCATCCAGACAT 1553
 400 PheAlaThrProLysMetGlnGluLeuPheGlyIleThrGlyS 414
 1554 ACACCCCTTGTACGCCCTCATGACGAGATGTTGTTGGCATCACAAGTA 1603
 414 et 414
 1604 GC 1605


```

317 MLeuLeuLeuGluProMetLeuLysPheHisTyrMetLeuLysLysLeuG 334
|||||
1127 ACTTCTACTGAGAGCCCATGCTGAATTCACATCACTGGAAGAAAGTGC 1176
334 LLeuHisGluGluGluTyrValLeuMetGlnAlaIleSerLeuPheSer 350
|||||
1177 AGCTGATGAGAGAGAGTATGCTGATGACAGCCATCCCTCTTCTGCC 1226
351 ProAspArgProGlyValLeuGlnHisArgValValAspGlnLeuGlnG 367
|||||
1227 CCAGACGCCGCCAGAGTGTCTGCAGACGCCGTGTGGACAGCTGAGGA 1276
367 uGlnPheAlaIlePheLeuLysSerTyrTleGluLysAsnArgProGlnP 384
|||||
1277 GCATTCGCCATTACTGCAAGTCTCACTGATGATGATGATGATGATGAT 1326
384 roAlaHisArgPheLeuPheLeuLysIleMetAlaMetLeuThrGlu... 399
|||||
1327 CTGCTCATAGTGTCTTGTCTGATGATGATGATGATGATGATGATGAT 1376
399 ..... 399
1377 CGCAGCATCATGCTGACGACACGCCGCTGCTGCGATCCAGGACAT 1426
400 ..... PheAlaThrProLeuMetGlnGluLeuPheGlyIleThrGlyS 414
1427 ACACCCCTTGTGACCCCTCATGAGAGTGTGTCGATCAGAGGTA 1476
414 er 414
1477 GC 1478

```

seq_name: /SIN08/4/cgdata/geneseq/geneseq/NA1999.DAT:AAx78808

seq_documentation_block:

ID AAX78808 standard; cDNA: 2850 BP.

AC AAX78808:

XX 06-SEP-1999 (first entry)

DE Human nNR7-1 cDNA.

XX nNR7; nNR7-1; nuclear trans-acting receptor protein; human; regulator;

KM Identification: downstream target gene; cell proliferation;

XX cell development; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 104..1525

ET /tag= a

ET /product: "nNR7-1"

XX MO9931129-A1.

XX 24-JUN-1999.

XX 11-DEC-1998; 98WO-0526364.

XX 14-OCT-1998; 98US-0104251.

XX 12-DEC-1997; 97US-0069401.

XX (MER1) MERCK & CO INC.

XX Chen F;

XX WPI; 1999-405024/34.

XX P-PSDB; AAY25411.

XX DNA encoding human nuclear receptors nNR7 and nNR7-1

PS Claim 28: Fig 4A-C; 80pp: English.

XX This invention describes the isolation of the novel human nuclear
 CC receptors nNR7 or nNR7-1. The nNR7 and nNR7-1 proteins are useful in
 CC the identification of downstream target genes and ligands regulating its
 CC activity. The nuclear receptor is involved in the regulation of in vivo
 CC cell proliferation and/or cell development. The nNR7 and nNR7-1
 CC polynucleotides, expression vectors and host cells are useful for the
 CC recombinant production of the protein.

SQ Sequence 2850 BP: 691 A; 736 C; 787 G; 636 T; 0 other;

alignment_scores:

Quality: 2167.00 Length: 434
 Ratio: 5.234 Gaps: 1
 Percent Similarity: 95.392 Percent Identity: 95.392

alignment_block:

US-09-276-935b-14 x AAX78808 ..

Align seq 1/1 to: AAX78808 from: 1 to: 2850

```

1 LeuGluValArgProLysGluSerTyrPheHisAlaAspPheValHisCy 17
|||||
221 CTGGAGGTGAGAGCCCAAGAAAGCTGGAACCATGCTGACTTTGTACACTG 270
17 sGluAspThrLeuSerValProGlyLysProSerValAsnAlaAspGluG 34
|||||
271 TCAGGACACAGAGTGTCTCTGGAAGCCAGTGTCAACGACAGATGAGG 320
34 LuValGlyGlyProGlnIleCysArgValGlyGlyAspLysAlaThrGly 50
|||||
321 AAGTGGAGGTCCCAAACTGCTGCTGATGAGGGAACACCCACTGCG 370
51 TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPhePheArgAr 67
|||||
371 TATCACTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
67 gAlaMetLysAlaAsnAlaArgLeuArgCysProPheArgLysGlyAlaG 84
|||||
421 GCCCATGAAACGCAAGCCCGCTGAGTGGCCCTTCCGGAAGGCGGCT 470
84 ySGluIleThrArgLysThrArgArgGlnGlyGlnAlaCysArgLeuArg 100
|||||
471 GCCAGATCACCCGGAAGACCCGCGACAGTGTCAAGCTGCGCGCTGCGC 520
101 LysCysLeuGluSerGlyMetLysLysGluMetIleMetSerAspGluAl 117
|||||
521 AAGTGGCTGAGAGAGCGGCATGAAGAGAGATGATGATGATGATGATGATG 570
117 aValGluGluArgArgAlaLeuIleLysArgLysLysSerLysArgThrG 134
|||||
571 CGTGGAGAGAGAGCGCGCTGATCAAGCGGAAGAAAGTGAACGAGACG 620
134 yThrGlnProLeuGlyValGlnGlyLeuThrGluGluGlnArgMetMet 150
|||||
621 GGACTCAGCCACTGGAGAGTGCAGGGGCTGACAGAGACACACGATGATG 670
151 IleArgGluLeuMetAspAlaGlnMetLysThrPheAspThrThrPheSe 167
|||||
671 ATCAGAGAGACTGATGACCGCTGATGATGATGATGATGATGATGATGATG 720
167 rHisPheLysAsnPheArgLeuProGlyValLeuSerSerLysGluLeu 184
|||||
721 CCATTTCAGAAATTCGCGCTCCGAGGGTCTGTCAGATGCTGCTGAGAT 776
184 euProGluSerLeuGlnAlaProSerArgGluGluAlaAlaLysTrpSer 200
|||||
771 TGCCAGAGTCTTCGAGGCCCATGAGGGAAGAAAGTGTGCCAAGTGGAGC 820
201 GlnValAlaGlyAspLeuLysSerLeuLysValSerLeuGlnLeuArgG 217
|||||

```


151	IIeAlGIuIeMeAspAlaGluMetIysThrPheAspIrrIrrIrrPheSe	167
725	ATCAGGAGGACTGATGAGCGCTCAGTGAACCTTTGACACTCCTTTC	774
167	ThIsPheIysAsnPheArGLeuProGlyValLeuSerSerGlyGluL	184
775	CCATTTCAGGAATTTCCGGCTCCGAGGGGTGTAGCGAGTGGTGGAGT	824
184	eupProGluSerLeuGlnAlaProSerArGgIuGluAlaAlaIysTrpSer	200
825	TCGCCAGACTCTCTCGAGGCCCATCGAGGAGGAAGAGCTGCCAAGTGGAGC	874
201	GlnValArGlyAspLeuGlySerLeuIysValSerLeuGlnLeuArG	217
875	CAGGTCCGGAAGATCTGTGCTCTTTGAAGTCTCTTCGACGCTCGGGG	924
217	YcLIuAspGlySerValTTPAsHTyLysProProAlaAspSerGlyL	234
925	CGAGAGTGCACATGTCGTGTGAACATCAACCCCGACCCGACAGTGGCGGA	974
234	YSGIuIThPheSerLeuLeuProHisMetAlaAspMetSerThyMet	250
975	AAGAGATCTTCTCCTGCTGCTGCCCATGCTGCTGACATGTCAACTACATG	1024
251	PheIysGlyIleIleSerPheAlaIysValIleSerTyPheArGAspLe	267
1025	TTCAAAAGGCATCATCAGCTTTCCTCAAGCATCATCTCTACTTCAGGACTT	1074
267	uProIThGlyuAspGlnIleSerLeuLeuIysGlyAlaAlaPheGluLeuG	284
1075	GCCCATTCAGGAGCCAGATCTCTCTCTCAAGGGGGCCCTTTCGAGCGGT	1124
284	YSGIuLeuArGPheAsnThrValPheAsnAlaGluThyGlyThrProGlu	300
1125	GTCAACTGATGATTTCAACACAGTGCTTCAACGGGAGACCTGGAACTGGAG	1174
301	CysGlyArGLeuSerTyrcysLeuGluAspThrAlaGlyGlyPheGlnG	317
1175	TGTGGCGGCTCTCTACTGCTTGAGAACACTGACGTGGCTTCACAGA	1224
317	IleuLeuLeuGluIupPrometLeuIysPheHisTyrcMetLeuIysLysLeu	334
1225	ACTTCTACTCGAGCCCATCTCTGAAATTCACACTACATGCTGAAGAGACTTC	1274
334	IleuHisGlnGluIuGlyrValIleuMetGlnAlaIleSerIupPheSer	350
1275	AGCTGCATGAGGAGAGTATGTGTGATGAGCAGCATCTCCCTTTTCC	1324
351	ProAspArGProGlyValLeuGlnHisArValValAspGlnLeuGlnG	367
1325	CCAGACCGCCCGAGGTGTGTGAGCAGCCCGGTGTGACACACTTCACAGA	1374
367	uGlnPheAlaIleThrLeuIysSerTyrlleGluIysAsnArGProGlnP	384
1375	GCAATTTGGCATTTACTCTGAGACTCTACATTGATTCATTCGAGCCCTCAGC	1424
384	roAlaHisArGpHeuPheLeuIysIleMetAlaMetLeuThrGln...	399
1425	CTGCTCATGTGTTCTGTCTCGAAGATCATGCTATGCTCACCGAGCTC	1474
399	399
1475	CGCAGCATCATGCTCAGCACACCGACCGGCTCTGCCCATCCAGGACAT	1524
400PheAlaIrrProIleuMetGlnIuLeuPheIysIleThrGlyS	414
1525	ACACGCCCTTGTACAGCCCTCATGAGAGASATGTTCCGACATCACAGTA	1574
414	er 414	
GC		
1575	GC 1576	

```

seq_name: /SI058/tcdata/genescq/genesequ/NA1599.DAT:AAK78807
seq_documentation_block:
ID   AAK78807 standard: cDNA: 3093 BP.
XX
XX   AAK78807:
XX
XX   06-SEP-1999 (first entry)
XX
XX   Human nNR7 cDNA.
XX
XX   nNR7: nNR7-1; nuclear trans-acting receptor protein; human; regulator;
XX   identification; downstream target gene; cell proliferation;
XX   cell development; ss.
XX
XX   Homo sapiens.
XX
XX   Key      Location/Qualifiers
XX   CDS       275..1676
XX   FT        /*tag= a
XX   FT        /product= "nNR7"
XX   FT        /note= "Partial sequence, no start codon given"
XX
XX   WO9931129-A1.
XX
XX   PD        24-JUN-1999.
XX
XX   PF        11-DEC-1998; 98WD-US26364.
XX
XX   PR        14-OCT-1998; 98US-0104251.
XX   PR        12-DEC-1997; 97US-0069401.
XX
XX   PA        (MER1 ) MERCK & CO INC.
XX
XX   PI        Chen F;
XX
XX   DR        WP1: 1999-405024/34.
XX   DR        P-PSDB: AAK75410.
XX
XX   PT        DNA encoding human nuclear receptors nNR7 and nNR7-1
XX
XX   PS        Claim 9; Fig 1A-C; 80pp; English.
XX
XX   CC        This invention describes the isolation of the novel human nuclear
XX   CC        receptors nNR7 or nNR7-1. The nNR7 and nNR7-1 proteins are useful in
XX   CC        the identification of downstream target genes and ligands regulation in
XX   CC        activity. The nuclear receptor is involved in the regulation of in vivo
XX   CC        cell proliferation and/or cell development. The nNR7 and nNR7-1
XX   CC        polynucleotides, expression vectors and host cells are useful for the
XX   CC        recombinant production of the protein.
XX
XX   Sequence 3093 BP: 774 A: 786 C: 843 G: 690 T: 0 other:
XX
XX
XX   alignment_scores:
XX       Quality: 2167.00      Length: 434
XX       Ratio: 5.234          Gaps: 1
XX   Percent Similarity: 95.392  Percent Identity: 95.392
XX
XX   alignment_block:
XX   US-09-276-935B-14 x AAK78807 ..
XX
XX   Align seq 1/1 to: AAK78807 from: 1 to: 3093
XX
XX   1 LengthVal:ATGProlysgluserTrpasnhtsAlaaspphValHisCy 17
XX   ||||||||||||||||||||||||||||||||||||||||||||||||
XX   372 CTGGAGCTAGACCCCAAGAAAGCTGGAAACCATGCTGACTTTGACACTG 421
XX
XX   17 scguasprhrgluserValProGlylysproservalasnaIaasplog 34
XX   ||||||||||||||||||||||||||||||||||||||||||||||||
XX   422 TAGAGACACAGAGTCTGTTCTGGAAGACCCAGTGTCAACGACACATAGG 471
XX
XX   34 lvalaGllylProGlnlleCySarValGlyscGlyAspIysAlaIthrcly 50

```



```

|||||
472 AATCTCAAGCTCTCCAAATCTGGCGGTATGTTGACCAACAGCTCACTGAGT 521
51 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 67
|||||
522 TATATATATATGATGAGATGTTGACCAAGTAAAGCTTTTTCAGGAG 571
67 GATATATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 84
|||||
572 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 621
84 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 100
|||||
622 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 671
101 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 117
|||||
672 AATCTCTTGGACAGCGGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721
117 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 134
|||||
722 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 771
134 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 150
|||||
772 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 821
151 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 167
|||||
822 AATCTCTTGGACAGCGGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871
167 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 184
|||||
872 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 921
184 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 200
|||||
922 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 971
201 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 217
|||||
972 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1021
217 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 234
|||||
1022 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1071
234 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 250
|||||
1072 AATCTCTTGGACAGCGGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1121
251 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 267
|||||
1122 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 1171
267 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 284
|||||
1172 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1221
284 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 300
|||||
1222 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1271
301 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 317
|||||
1272 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 1321
317 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 344
|||||
1322 AATCTCTTGGACAGCGGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
344 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 350
|||||

```

```

1372 AATCTCTTGGACAGCGGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1421
351 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 367
|||||
1422 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 1471
367 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 384
|||||
1472 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1521
384 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 399
|||||
1522 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 1571
399 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 399
|||||
1572 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1621
400 TTTTATATGASNAValMetThrCysGluGlyCysGluSerGlyThrGluGluVal 414
|||||
1622 ACACGCTTTGCTACGCGGCTCATGACAGAGATTGTCGACATCAAGAGTA 1671
414 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
|||||
1672 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
|||||
1672 GATATATATGAGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
|||||

```

seq_name: /SUSB/494data/geneseq/geneseq/NA1999.DAT:AAK59974

seq_documentation_block:

ID: AAK59974 standard: DNA: 3243 BP.

AAK59974:

04-AUG-1999 (first entry)

DNM encoding an intranuclear receptor protein.

Human: intranuclear receptor protein; drug development; diagnosis;

treatment; SS.

Human sapiens.

J01127872-A.

18-MAY-1999.

07-AUG-1998: 98JP-0224172.

11-AUG-1997: 97JP-0230335.

(NIST) JAPAN TOBACCO INC.

WPI: 1999-350340/40.

P-PSDB: AAK15946.

New intranuclear receptor protein - useful for drug development and

diagnosis and treatment of disease.

Claim 10: Page 32-35; 38pp; Japanese.

The present sequence encodes a human intranuclear receptor protein.

The nucleotide acid sequence was isolated from a human adult cDNA

library using a Swedish AN023 derived probe. The protein can

be used for the development of drugs and diagnosis and treatment

of various diseases.

Sequence 3243 BP: 818 A; 806 C; 862 G; 737 T; 0 other;

alignment_scores: length: 424

quality: 2167.00 gaps: 1

Percent Similarity: 95.392 Percent Identity: 95.392

alignment block:

US-09-276-935b-14 x AAX59974

Align seq 1/1 to: AAX59974 from: 1 to: 3243

```

1 LeuGIValArpProLysGluSerTrpPasnHisAlaAspPheValHisGly 17
461 CTGGAGGTAGACCCAAACAAAGCTGGAAACCATGCTGACTTTGTACACTG 510
17 sGIuAspThrGluSerValProGlyLysPheSerValAsnAlaAspGluG 34
511 TGGAGCACAGAGTGTCTGCGAAAGCCAGTGTCAAGCGCAGATGAGG 560
34 LuValGlyGlyProGlnIleCysArgValCysGlyAspLysAlaThrGly 50
561 AAGTCGGAGGCTCCCAAAATCTGCGGTATGTGGGACCAAGGCCACTGGC 610
51 TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPhePheArg 67
611 TATCACTTAATGTCAATGATGTCAAGATGTCAAGGCTTTTTCAGGAG 660
67 gAluMetLysArgAsnAlaArgLeuArgCysProPheArgLysGlyAlaG 84
661 GGCCATGAAGACGACAGCCCGGCTGAGTGCCCTTCGGAAAGGCGCCT 710
84 ysgLulleThrArgLysThrArgArgGlnCysGlnAlaCysArgLeuArg 100
711 GGCGAGTCACCCGGAAGACCCGCGCAGACTGCCAGGCTGCGCCCTGGCG 760
101 LysCysLeuGluSerGlyMetLysLysGluMetLleMetSerAspGlu 117
761 AAGTGGCTGGACAGCCGCGCATGAAGAGAGATGATGATGCTCCAGAGGC 810
117 aValGluIuArgArgAlaLeuIleLysArgLysLysSerGluArgThrG 134
811 CCGAGAGAGAGAGCGGCGCTTATCAAGCGGAAGAAAGTGAAGCGACAG 860
134 lYThrGlnProLeuGlyValGlnGlyLeuThrGlnGlnArgMetMet 150
861 GGACTGAGCCACTGGGAGTGCAGGGGCTACACAGAGAGCAGCGGATGATG 910
151 lIeArgGluLeuMetAspAlaGlnMetLysThrPheAspThrThrPhe 167
911 ATCAGGAGAGTGTGACCGCTGAGATGAACCTTGTACACTACCTTCTG 960
167 rHisPheLysAsnPheArgLeuProGlyValLeuSerSerGlyCysGlu 184
961 CCATTTCAGAAATTTCGCGCTGCCAGGGGTCTTACGACGTGGCGAGT 1010
184 euProGluSerLeuGlnAlaProSerArgGlnGlnAlaAlaLysTrpSer 200
1011 TGGCAGAGCTCTGACAGCCCATCAAGCAACACCTGCAAGTGGAGAC 1060
201 GluValArgLysAspLeuCysSerLeuLysValSerLeuGlnLeuArg 217
1061 CAGGTCGGGAAGATCTGTGCTCTTGAAGGTCTCTGTGAGGTGGGGG 1110
217 yGIuAspGlySerValTrpAsnTrpLysProPheAlaAspSerGlyGly 234
1111 GAGAGATGGAGTGTCTGGAACATACAAACCCGAGCGCAGAGTGGGGGA 1160
234 ysgLullePheSerLeuLeuProHisMetAlaAspMetSerThrTrpMet 250
1161 AAGAGATCTTCTCCCTGCTGCCACATGCTACATGTCACCTCATG 1210
251 PheLysGlyLleIleSerPheAlaLysValIleSerTrpThrArgAsp 267
1211 TTCAAAGGCATCATCAGCTTGGCAAGATATCTCTACTTCAGGCACTT 1260
267 uProIleGluAspGlnIleSerLeuLeuLysGlyAlaAlaPheGluLeu 284
1261 GCCATTCGAGACCAATCTCCTGCTGAAGGGGCGCGTTTCAGACTGT 1310

```

```

284 ysgLulleuArgPheAsnThrValPheAsnAlaGluThrGlyThrTrpGlu 300
1311 GTCAACTGAGATTCAACACAGTGTTCACCGCAGACTGGAATCTGGAG 1346
301 CysGlyArgLeuSerTrpCysLeuGluAspPheAlaGlyGlyPheGluG 317
1361 TGTGGTCGGCTGTCTCTACTGCTTGGAAACACTGAGAGTGGCTTCAGCA 1410
317 lLeuLeuLeuGluProMetLeuLysPheHisTyrMetLeuLysLysGlu 334
1411 ACTTCTACGAGGCCCATGCTGAATTCACACTACATGCTGAAGAAAGTGC 1460
334 lLeuHisGluGlnGluTyrValLeuMetGlnAlaIleSerLeuPheSer 350
1461 AACTCATCAGAGAGATGTGCTGATGCACAGCCCATCTCCCTTCTTGC 1510
351 ProAspArgProGlyValLeuGlnHisArgValAlaAspGlnLeuGlnG 367
1511 CCAGACCGTCCAGGTGTGCTGACGACCGCGTGTGACACCACTGACAGA 1566
367 uGlnPheAlaIleThrLeuLysSerTrpIleGlnCysAsnArgProGln 384
1561 GCAATTCGGCATTACTCTGAAGTCTCTACATTGAATGCAATGGCCCGCAG 1610
384 roAlaHisArgPheLeuPheLeuLysIleMetAlaMetLeuThrGlu... 399
1611 CTGCTCATAGAGTCTTCTGCTCGAAGATCATGCTATCTCTACCGAGCTC 1660
399 ..... 399
1661 CGCAGCATCAATGCTCAGCAGCCGCGCTGCTGCGCATCCAGAGACAT 1710
400 .....PheAlaIleProLeuMetGlnLeuPheGlyIleThrLys 414
1711 ACACCGCTTCTGCTACGCGCCCTCATGAGAGATTGTCGAGATCAGAGG 1760
414 et 414
1761 GC 1762

seq_name: /SIDS8/4/udata/geneseq/geneseqr/NA1999.DAT.AAX56242
seq_documentation_block:
ID: AAX56242 standard: cDNA: 2910 BP.
XX
AC: AAX56242:
XX
DT: 16-JUL-1999 (first entry)
XX
DE: Human vitamin D receptor related gamma protein encoding cDNA.
XX
KW: Human; vitamin D receptor related protein; VDR; obesity; diabetes;
anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;
hypercholesterolemia; hyperlipoproteinaemia; osteoporosis; tumour;
hyperproliferative skin disorder; hyperthyroidism; ss.
XX
OS: Homo sapiens.
XX
PN: W09919354-A1.
XX
PD: 22-APR-1999.
XX
PF: 31-AUG-1998: 98WD-SE01548.
XX
PR: 31-MAR-1998: 98SE-0001148.
PR: 14-OCT-1997: 97SE-0003745.
XX
FA: (PNUA) PHARMACIA & UPJOHN AB.
XX
PI: Berkenstam A, Dahlberg M:
XX
WP: 1999-302508/25.

```


108 P (SDS): MAY0515.

XX New vitamin D receptor related (VDR) polypeptides, useful for
 109 treating obesity, diabetes, anorexia and rheumatoid arthritis
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

seq_name: /SIDS8/gcdata/geneseq/geneseq/NA199.DAT:AAx89080

seq_documentation_block:

ID AAX89080 standard: DNA: 2068 bp.

AC AAX89080:

DT 14-SEP-1999 (first entry)

DE Human steroid and xenobiotic receptor (SXR) encoding DNA.

KM Nuclear receptor: SXR: steroid and xenobiotic receptor; RXR: human; retinoid X receptor; P450 gene; steroid hormone; steroid metabolism; phytoestrogen; calcium-channel blocker; steroid toxicity; tuberculosis; breast cancer; osteoporosis; Cushing syndrome; vitiligo; hirsutism; polycystic ovarian disease; cancer; colorectal; prostatic; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 583..1887

FT /tag= a

FT /transl_except= (pos:1261..1263, aa:Xaa)

FT /note= "Xaa= unknown; the start codon is not indicated"

XX W09355246.A1.

XX 15-JUL-1999.

XX 08-JAN-1999; 99WO-US00490.

XX 09-JAN-1998; 98US-0005286.

XX (SALK) SALK INST BIOLOGICAL STUDIES.

XX Blumberg B, Evans RM;

XX WPI: 1999-419349/75.

XX P-PSDB: AAT21799.

PT New steroid and xenobiotic receptor, used to identify modulators for controlling metabolism of steroids and xenobiotics, e.g. reducing their toxicity

PS Claim 10; Fig 1A; 83bp; English.

CC The invention relates to a novel nuclear receptor polypeptide, designated SXR (steroid and xenobiotic receptor). SXR (i) forms a heterodimer with retinoid X receptor (RXR), (ii) binds to a direct or inverted repeat response element motif based on the half-site AGTTCA, (iii) activates transcription through response elements present in steroid-inducible P450 genes, in response to a wide variety of natural and synthetic steroid hormones and (iv) is prominently expressed in liver and intestine. SXR regulates expression of catabolic enzymes, in response to many different steroids, and thus affects metabolism. SXR is a broad specificity, low-affinity receptor for reducing excessive levels of steroids in the circulation. (Ant)agonists of SXR are used to regulate metabolism of steroids particularly phytoestrogens or calcium-channel blockers, to reduce steroid toxicity in subjects being treated with steroids, e.g. in cases of tuberculosis (treated with rifampin and related compounds), breast cancer (treated with tamoxifen, raloxifene etc.) or osteoporosis (treated with Vitamin K), or to slow metabolism of therapeutic steroids. Also, modulating endogenous SXR is used to treat disease, particularly an agonist is used where endogenous steroid levels are excessive (e.g. Cushing syndrome; vitiligo and hirsutism in women; polycystic ovarian disease; β 17- or 21-hydroxylase deficiency; 3 β -hydroxysteroid dehydrogenase deficiency, or breast, colorectal or prostatic cancer), while antagonists are used where endogenous steroid levels are too low. Cells that express SXR are used to identify compounds likely to be involved in undesirable drug interactions. Antibodies specific for SXR are used in immunohistochemical testing for studying distribution/ expression density of SXR, also for diagnosis and therapeutically as

CC antagonist. The present sequence represents the longest SXR cDNA clone encoding the SXR polypeptide.

XX Sequence 2068 bp: 520 A; 541 C; 586 G; 420 T; 1 other;

alignment_scores:

Quality: 2112.00 Length: 435
Ratio: 5.126 Gaps: 2
Percent Similarity: 94.71% Percent Identity: 94.71%

alignment_block:

US-09-276-935b-14 x AAX89080 ..

Align seq 1/1 to: AAX89080 from: 1 to: 2068

1 LeuGluValArgProLysGluSerTrpAsnHisAlaAspPheValHisGly 17
|||||
583 CTGACAGTACAGCCCAAGAAAGAGCTGCAACCATCTGACTTTGACACTG 632
|||||
17 sGluAspThrGluSerValProGlyLysProSerValAsnAlaAspGluG 34
|||||
633 TGAGACACACAGATCTGCTCTGAAAGCCAGTGTCAAGCAGACATGAG 682
|||||
34 lValGlyGlyProGluHisGlySerValGlyGlyAspLysAlaThrGly 50
|||||
683 AAGTCGACAGTCCCAATCTCCCTGTATGTGGGACAAAGCCACATGAC 732
|||||
51 TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPhePheArg 67
|||||
733 TATCACTTCATGTCATGACATGATGAGATGCAAGGCTTTTCAGAGAG 782
|||||
67 gAlaMetLysArgAsnAlaArgLeuArgCysProPheArgLysGlyAlaC 84
|||||
783 GGCATGAAACCAACCCCGGCTGAGTCCCTTCGGAAGGCGGCTCT 832
|||||
84 YsGluHisThrArgLysThrArgArgGlnCysGluAlaCysArgLeuArg 100
|||||
833 GCGAGATCACCGGAAAGACCGCGGACAGTGCAGGCGCTGCGCTTGGC 882
|||||
101 LysCysLeuGluSerGlyMetLysLysGluMetIleMetSerAspGluAl 117
|||||
883 AAGTGGCTGAGACAGCGGCTGAAAGAGATGATCATGTCCGACGAGGCG 932
|||||
117 dValGluGluArgValAlaLeuIleLysArgLysLysSerGluArgThr 134
|||||
933 CCGTAGAGAGAGCGCGGCTGATCAAGCGGAAGAAATGCAACCGACAG 982
|||||
134 lYThrGlnProLeuGlyValGlnGlyLeuThrGluGluArgMetMet 150
|||||
983 GACTCTACGCTCTGGAGTGCAGCGGCTGACAGAGAGAGCGATGATG 1032
|||||
151 lIleArgIleMetLeuAspAlaGlnMetLysThrPheAspThrThrPhe 167
|||||
1033 ATCAGGAGACTGATGAGCGCTCAGATGAAAGCTTTGACACTCTTCG 1082
|||||
1083 rHisPheLysAsnPheArgLeuProGlyValLeuSerSerGlyCysGlu 184
|||||
184 eProGluSerLeuGlnAlaProSerArgGluGluAlaAlaLysTrpPro 200
|||||
1133 TCCGACAGCTCTGCGAGGCCCATCGAGGGAAGAACTGCTCAAGTGAC 1182
|||||
201 GluValArgLysAspLeuLysSerLeuLysValSerLeuGlnIleuVal 217
|||||
1183 CAGGTCGCGAAGATGTGTCTTTGAAGGTCTCTCTCAAGCTGCGGG 1232
|||||
217 LysLysAspLysSerValTrpAsnTrpLysProProAlaAspSerGlyGly 233
|||||
1233 GGGAGGATGGCAATGTCTGCAACTACAAACCCGACCCGACATGACGCA 1282
|||||
234 LysGluHisPheSerLeuLeuProHisMetAlaAspMetSerThrTyre 250


```

553 GGGCCCCACAGCCCGAGGAGC...AGCTGACGTCCAGCGACGACGACACT 599
150 etlleargluleuMeCaspalaIgmellsthrpheasphrThrPhe 166
600 TTCTGACTGACACTGCTGGGCCCGACACCAAACTTGTACTTCAACTTC 649
167 SerHisPheLysaspheargleuProgllyValLeuSerSerGlycylsrl 183
650 ACCTTCTCCAGAGAACTTTCG... 670
183 uleuProgluSerleuGlnAlaProserArggluGlnAlaAlaIstTps 200
671 .....CCATTAAGA..... 679
200 erGlnValArglyaspLeuGlySerleuIstValSerleuGlnleuArg 216
680 .....AGA 682
217 GlygluaspGlySerValTrrpasnTrrlyProProAlaaspSerGly 233
683 TCTTCAGACCCAACT.....CAGAGCCCGCAAGCCACCTCT..... 718
233 ylysgIuilepheSerleuProHisMetAlaaspMetSerThrTyrM 250
719 .TCTGAAGCCTTTTGTGCTGACTCATATGCTGACCTGCTTACTTACA 767
250 ePheLysGlyIleIleSerPheAlaIstValIleSerTyrPheargasp 266
768 TGATCAAGGGCAGATCAGCTTGGCCAAAATGCTCCCTCACTTCAAGAG 817
267 leuProIleGluaspGlnIleSerleuLeuLysGlyAlaAlaPheGlu 283
818 CTGAGACTTGAAGCAAAATGCTCTCTGCTGAAAGTTTGTACCGAGGT 867
283 uCysGlnleuArgPheasnThrValPheasnIagIurhGlyThrTyrP 300
868 TTCTGTGATCGCATTCACACTGCTGTTTAACTGTGACCAACATAGCTG 917
300 IuGlyGlyArgLeuSerTyrCysLeuGluaspThr...AlaGlyGlyPhe 315
918 AGTGTGGCCCTTTCACCTTGAACAGTGAAGATATGTTCTGAGCCCTTC 967
316 GlnGlnleuLeuGluProMetLeuLysPheHisTyrMetleuLysI 332
968 CGTCACTGCTTCTGAGACCCCTGGTGAGATTCATCGCATGATGAGACA 1017
332 sleuGlnleuHisGluGluIuTyrValleuMetGlnAlaIleSerleu 149
1018 ACTGATGTACAGAGTGAAGATATGCCCATGATGCGCGCTGTGCATT 1067
349 heSerProAsparTroglyValleuGlnHisArgValAlaaspGlnleu 365
1068 TCCCTCTTACACCGCCGCTGTGCGACTGAGAGATGAGAGAGCTG 1117
366 GlnGlnleuPheAlaIleThrleuLysSerTyrIleGluCysasnArg 381
1118 CAGAGACACTTGGCCCTGAGACTAAAGATTTGATGACAGCGACAGGCC 1167
382 .ProGlnProAlaHisArgPheleuPheLysIleMetAlaMetleu 398
1168 CCCCTCCCGCAGACAGCGCTCTCTTACACCAATCATGATGTGTCTGA 1217
398 hrcIuPhe..... 400
1218 CAGAGCTTCCGACACTCAATGACATACAGACAGACAGCTCTGGAGATC 1267
401 .....AlaThrProleuMetGlnGluIuPheGly 410
1268 TGGGACATTCAGCCTGATGCCACCCCACTTATGCGAGAAATCTTTTGA 1315
seq_name: /SID8/gcdata/geneseq/geneseqn/NA198.DAT:AAV03129

```

```

seq_documentation_block:
ID      AAV03129 standard; cDNA; 1960 BP.
XX
AC      AAV03129;
XX
DT      11-JUN-1998 (first entry)
XX
DE      cDNA encoding rat vitamin D receptor (VDR).
XX
KW      Rat; vitamin D receptor; isoform protein; VDR1; VDR0; diagnosis;
KW      dominant negative receptor; signal transmission channel;
KW      bone density disorder; screening; ds.
XX
OS      Rattus rattus.
XX
FH      Key
FH      Location/Qualifiers
FT      CDS
FT      12..1283
FT      /*tag= a
FT      /transl_except= (pos:1089..1091, aa:Gly)
XX
PN      W09747172-A1.
XX
PD      18-Dec-1997.
XX
PE      10-JUN-1997; 97W0-1E00947.
XX
PF      10-JUN-1996; 96JP-0194179.
XX
PR      (CHUS ) CHUGAI PHARM CO LTD.
PA      (CHUS ) CHUGAI SEIYAKU KK.
XX
PI      Kato S, Ueno K;
XX
DR      WPI: 1998-051917/05.
DR      P-PSDB: AAW47509.
XX
PT      DNA encoding a vitamin D receptor isoform protein - useful for bone
PT      density determination and for screening substances for vitamin D
PT      activity
XX
PS      Disclosure: Fig 1: 46pp; Japanese.
XX
CC      A novel cDNA sequence encodes the rat vitamin D receptor isoform
CC      protein (VDR1). The isoform differs from the normal receptor
CC      (VDR0), which is encoded by the present sequence, in having the
CC      vitamin D response element curtailed by 86 residues, and having an
CC      extra 19 residues inserted at the C-terminal of this element. It
CC      acts as a dominant negative receptor in the vitamin D signal
CC      transmission channel.
CC      The isoform protein can be used to diagnose bone density disorders,
CC      and screen for substances having potential vitamin D-like activity.
XX
SQ      Sequence 1960 BP: 432 A; 620 C; 497 G; 411 T; 0 other;
XX
alignment_scores:
Quality: 762.00 Length: 410
Ratio: 2.712 Gaps: 8
Percent Similarity: 68.537 Percent Identity: 40.000
alignment_block:
US-09-276-935b-14 x AAV03129 ..
Align seg 1/1 to: AAV03129 from: 1 to: 1960
38 ProGlnIleCysArgValCysGlyAspLysAlaThrGlyThrHisPheAs 54
|||||  |||||||  |||||||  |||||||  |||||||  |||||||
72 CCGCGATGTGTGAGATGTGTGTGAGAGCGACGACAGCGCTTCCATTTCA 121
54 nValMetThrCysGluGlyCysLysGlyPhePheArgArgAlaMetLysA 71
|||||  |||||||  |||||||  |||||||  |||||||  |||||||
122 TCTATGACCTGTGAAGGCTGCAAGGTTTCTTCAAGCGCGACATGAAAGC 171

```


Quality: 762.00 Length: 410
 Ratio: 2.712 Gaps: 8
 Percent Similarity: 68.537 Percent Identity: 40.000

alignment_block:
 us-09-276-935b-14 x AA051425 ..

Align seg 1/1 to: AA051425 from: 1 to: 2043

```

38 ProGlnIleCysArgValCysGlyAspLysAlaThrGlyTyrHisPheAs 54
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
155 CCGCGATCTGTGAGTGTGTGACAGCCAGACAGGCTTCCACTTCA 204
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
54 nValMetThrCysGlyGlyCysLysGlyPhePheArgAlaMetLysA 71
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
205 TGCTATGACCTGTGAAGCTGCAAGGTTCTTCAGCCGACCATGAAC 254
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
71 rGAsnAlaArgLeuArgCysProPheArgLysGlyAlaCysGluIleThr 87
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
255 GGAAGGCCCTGTTCACTCTCCCTTC...AATGAGATTGCCGCTACGC 301
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
88 ArgLysThrArgArgGlnCysGlnAlaCysArgLeuArgLysCysLeuG 104
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
302 AAGGACACACCGGCGACACTGCGAGGCTCCGCGCTCAACGCTGTGTGA 351
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
104 nSerGlyMetLysLysGluMetIleMetSerAspGluAlaValGluGlu 121
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
352 CATCGCATGATGAAGAGTTCATCTCTACAGATAGAGGATACAGCTGA 401
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
121 rArgAlaLeuIleLysArgLysLysSerGluArgThrGlyThrGlnPro 137
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
402 AGAGGAGCATGATATGAGAGAGAAAGAGAGAGGCTTGAAGGACACT 451
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
138 LeuGlyValGlnGlyLeuThrGlnGluGlnArgMetMetIleArgGlu 154
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
452 CTGAGGCCCAAG...CTATCTGAGAGAGAGACAGACATCATGACCATCT 498
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
154 nMetAspAlaGlnMetLysThrPheAspThrThrPheSerHisPheLys 171
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
499 GCTGAGCGCCACACACAGACCTATACCCGACCTACGCTCTCAGG 548
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
171 snPheArgLeuProGlyValLeu.....SerSerGlyCysLeuPro 185
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
549 ACTTCGGGCTCCAGCTTGTATGAGAGAGATACAGAGGACTATTCTGCA 598
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
186 GluSerLeuAlaIleProSerArgGlnGluAlaAlaLysTrpSerLys 202
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
599 AGC.....CCACACTCAGCTTCTCGGGAACCTCTCTCTCTC 636
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
202 IArgLysAspLeuCysSerLeuLysVal..... 211
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
637 GAGCTCTGACCTGACACCCACCTACATGACATGATGAGACCATCGGCT 686
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
212 .....SerLeuGlnLeuArgGlyLysGlyLysPheSerValTrpAsnTrpLys 226
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
687 TTTCACCTGATCTGAGAGCGAGAGATTCTCAT..... 721
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
227 ProProAlaAspSerGlyLysGlyLysGluIlePheSerLeuLeuProHis 243
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
722 GACCCCTGTGACTCTGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 771
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
243 rAlaAspMetSerThrTyrMetPheLysGlyLysIleLeuSerPheAlaLys 260
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
772 GGGTGAACCTTGATGATTCAGCATCCAAAAAGGTCATCGGCTTTCGCA 821
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
260 aIleSerTyrPheArgAspLeuProIleGluAspGlnIleSerLeuLeu 276
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
822 TGATCCAGGATTCAGGATCTTACTCGGATGACAGATTCGCTCTT 871
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
277 LysGlyAlaAlaPheGluLeuCysGlnLeuArgPheAsnThrValPheAs 293
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
872 AAGTCAGGCCCATTCAGGTGATCATGTTACGCTCCACACCATCTTTCAG 921
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

```

293 nAlaGluThrGlyThrTrpGluCysGlyArgLeuSerTyrCysLeuGlu 310
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
922 CATGATGATATGTCTGTGAGTGTGTGACGCGAGAGACTACAGTACAGG 971
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
310 spThr.....AlaGlyGlyPheGlnGlnLeuLeuLeuGluProMet 323
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
972 TCACGATGATGTCGCAAGCTGTGGCAGACACCTTGAGCTGATGATGCTGT 1021
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
324 LeuLysPheHisTyrMetLeuLysLysLeuGlnLeuHisGluGluTyr 340
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1022 ATAAAGTTCAGAGTGGGCTGTGAAGAGCTGAATTACATGAGCAAGACA 1072
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
340 rValLeuMetGlnAlaIleSerLeuPheSerProAspArgProGlyVal 357
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1072 TGTCCTTCTCAAGCATGCTCATCTGCTCTCTCCGACCCGACCTGCGCTC 1121
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
357 euGlnHisArgValValAspGlnLeuGlnGlnGlnIlePheAlaIleThr 373
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1122 AGAGCCGCAAGCTGTGTGAGCCATTCAGAGACCGGCTATCCAGACAGG 1171
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
374 LysSerTyrIleGluCysAsnArgProGlnProAlaHisArgPheLeu 390
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1172 CAGACCTACATCGCTGCCGCCACCCGCCGACGACGACGACGCTCTA 1221
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
390 eLeuLysIleMetAlaMetLeuThrGluPhe..... 400
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1222 TGCCAGATGATCCAGAAACTGGCGACCTGGAGGCTCAACAGGAGAC 1271
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
400 ..... 400
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1272 ACTCAACAATAACGCTCCCTCTCTCTCCAGCCGAGACATAGCATAG 1321
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
401 AlaThrProLeuMetGlnGluLeuPheGly 410
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
1322 CTCACACCCCTTGTCTGAGGCTGTTCGCGC 1351
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
seq_name: /SIUS8/gcdata/geneseq/geneseq/NA1999.DAT:AAx14789
seq_documentation_block:
ID AAX34789 standard: DNA: 1382 BP.
XX
AC AAX34789;
XX
DT 06-JUL-1999 (first entry)
XX
DE Nucleotide sequence of human vitamin D receptor (VDR) gene transcript 9.
XX
KW Vitamin D receptor; VDR; hvDR; variant; isoform; SRC-1; GRIP-1; TFIIB;
KW cofactor; human; ds.
XX
OS Homo sapiens.
XX
PN M0916872-A1.
XX
FD 08-APR-1999.
XX
PE 29-SEP-1998: 98WD-AU00817.
XX
PR 29-SEP-1997: 97AU-0009500.
XX
PA (GARV-) GARVAN INST MEDICAL RES.
XX
PI Crofts LA, Eisman JA, Hancock MS, Morrison NA.
XX
DR P-PSDB: AAY09036.
XX
WP1: 1999-263693/22.
XX
DR P-PSDB: AAY09036.
XX
PT New polynucleotides which encode novel isoforms of the human vitamin
PT D receptor or variant transcripts for hvDR
XX
PS Claim 4: Fig 6: 56pp; English.
XX
CC The invention relates to isolated polynucleotides which encode novel

```


Thu Jul 12 14:23:34 2001

us-09-276-935b-14.p2n.rng

Page 19

1
2
3
4

5

OM of: US-09-276-935b-14 to: EST: * out_format: p1s

Date: Jul 11, 2001 5:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ p2n_model -DEV=klp
-O=/cgn2/US09276935/rmatc_11072001_091607_14534/app_query.fasta.1.480
-DB=EST -QMT=fastlap -SUFFIX=p2n.first -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOPCP=4.500
-GAPPEXT=0.050 -XGAPOP=10.000 -XGAPPEXT=0.500 -PCAPCP=6.000
-DELEXT=7.000 -YCAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-LIST=45 -DOCALLIGN=200 -THR_SCORE=perc -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTENT=p1s -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09276935 -CGN1_1_3127 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THRSADS=1

Search information block:

Query: US-09-276-935b-14
Query length: 414
Database: EST: *
Database sequences: 10228115
Database length: 431459454
Search time (sec): 1360.000000

Score list:

Sequence	Strd Orig	ZScore	EScore Len	Documentation
gb_hnc:AK018630	+	1684.50	3265.29	9.3e-173 2525
gb_est124:AT146915	+	610.50	1172.86	4.3e-56 794
gb_est17:AT148656	+	595.00	1152.38	4.5e-55 343
gb_hnc:AK011108	+	545.50	1047.28	3.2e-49 677
gb_est31:AU140214	+	522.00	998.67	1.6e-46 837
gb_est52:AM671811	+	451.50	866.70	3.7e-39 493
gb_hnc:AK002513	+	399.50	749.91	1.2e-32 1728
gb_est189:BF57692	+	396.00	753.91	7.1e-33 659
gb_est143:AM158294	+	389.50	746.68	1.8e-32 432
gb_est176:BE573656	+	387.50	738.17	5.4e-32 637
gb_est42:AM107515	+	383.00	723.41	3.6e-31 1053
gb_est24:AT168052	+	372.00	708.46	2.4e-30 601
gb_est183:BF102347	+	369.50	704.39	4.1e-30 560
gb_est191:BF686868	+	369.00	701.58	5.8e-30 654
gb_est10:AA679591	+	358.00	694.00	1.5e-29 200
gb_est92:BF788461	+	356.50	676.66	1.4e-28 677
gb_est15:AT049299	+	348.50	657.90	1.6e-27 878
gb_est83:BF080344	+	345.50	657.33	1.7e-27 560
gb_est198:BF681373	+	340.50	642.24	1.2e-26 876
gb_est129:AL555330	+	338.50	636.85	2.4e-26 992
gb_est41:AM012632	+	336.50	640.11	1.6e-26 540
gb_est148:AM511148	+	336.00	640.95	1.4e-26 463
gb_est15:AT049299	+	335.50	639.97	1.6e-26 463
gb_est175:BE542994	+	320.50	601.18	2.3e-24 1024
gb_est116:AT080863	+	314.50	596.99	3.9e-24 539
gb_est129:AL555330	+	306.50	579.96	3.5e-23 603
gb_est16:AT133241	+	303.50	571.76	1.0e-22 735
gb_est80:BE871201	+	303.50	569.43	1.3e-22 895
gb_est129:AL555330	+	303.50	565.12	2.3e-22 851
gb_est175:BE542994	+	300.50	568.46	3.5e-22 589
gb_est97:BF176463	+	300.50	561.22	3.8e-22 1090
gb_est33:AAV010557	+	297.00	544.74	3.7e-21 215
gb_hnc:AK004709	+	297.00	547.39	3.2e-21 248
gb_est44:AM211123	+	290.50	547.39	2.3e-21 726
gb_est72:BE233303	+	286.50	543.78	3.6e-21 467
gb_est28:AL551072	+	286.00	540.40	5.4e-21 563
gb_est74:HE410116	+	285.00	530.66	1.9e-20 1105
gb_est129:AL558946	+	282.00	527.32	3.0e-20 891

gb_est40:AV655930 + 280.50 528.23 2.6e-20 643 + AV655930 AV655930 GLC Homo S
gb_est74:BE390042 + 280.50 527.10 3.1e-20 708 + BE390042 601295233F1 NIH_MGC
gb_est40:AV652750 + 279.00 541.29 1.8e-20 387 + AV652750 AV652750 GLC Homo S
gb_est40:AV652653 + 279.00 531.11 1.8e-20 393 + AV652653 AV652653 GLC Homo S

seq_name: gb_hnc:AK018630

seq documentation block:

LOCUS AK018630 2525 bp mRNA HNC US-FR-2001
DEFINITION Mus musculus adult male cecum cDNA, RIKEN full-length enriched library, clone:9130016118, full insert sequence.
ACCESSION AK018630
VERSION AK018630.1 GI:12858434
KEYWORDS CAP trapper.
SOURCE Mus musculus (Strain:C57BL/6) adult male cecum cDNA to mRNA, clone:9130016118, RIKEN full-length enriched mouse cDNA library clone:9130016118.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathu; Muridae; Murinae; Mus.

REFERENCE

1 (sites)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Methods Enzymol. 303, 19-44 (1999)

REFERENCE

2 (sites)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Morimatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (11), 1617-1630 (2000)

TITLE

JOURNAL MEDLINE
Genome Res. 10 (11), 1617-1630 (2000)
20499374

REFERENCE

3 (sites)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, D., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Kimuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishie, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikeda, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Toyama, Y., Izawa, M., Ohara, E., Matsui, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Okazaki, Y., Morimatsu, M., Inoue, Y. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE

20530913
4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE

JOURNAL MEDLINE
Nature 409, 685-690 (2001)
5 (bases 1 to 2525)

REFERENCE

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arikawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hara, K., Hara, A., Hayatsu, N., Hiramoto, K., Hirata, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Nomazaki, K., Ohno, M., Okazaki, Y., Okada, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanabe, Y., Tejima, Y., Toyama, Y., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL MEDLINE
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genome Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, url: http://genome-gsc.riken.go.jp/, Tel: 81 45-503-9222, Fax: 81-45-503-9216)

COMMENT

Please visit our web site (http://genome-gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.


```

399 ..... 399
1527 GCAAGCATCAGCCGACGCAACCCAGCAGTGTCTGCCATCCAAAGACTC 1576
400 .....PhealathrProleumEtGIndleupheglyleThr 412
1577 GCACCCCTTGGCACCCTCATGCAAGACTTATTACGACGACA 1622

seq_name: gb_est24:A1746915

seq_documentation_block: 794 bp mRNA EST 22-JUN-1999
LOCUS A1746915
DEFINITION U109a10.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
IMAGE:2076282 5' similar to U109a10.y1 054915 054915 PRESNAME X RECEPTOR.
; mRNA sequence.
ACCESSION A1746915
VERSION A1746915.1 GI:5125179
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 794)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person
,B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Riltter
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The Mashu-NCI Mouse EST Project 1999
Unpublished (1999)
CONTACT: Marra M/Washu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@watson.wustl.edu
This clone is available royalty-free through LNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:994778
Possible reversed clone: similarity on wrong strand
Seq primer: custom primer used
High quality sequence stop: 472.
FEATURES
Source
location/Qualifiers
1..794
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:2076282"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGGCCCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTCTGCTCTAAAGCTGGC and 3' end primer
CGACCTGCACCTGAGCACA."
BASE COUNT 206 a 180 c 253 g 152 t 3 others
ORIGIN
alignment_scores:
Quality: 610.50 Length: 156
Ratio: 4.424 Gaps: 1
Percent Similarity: 88.462 Percent Identity: 74.359

```

```

alignment_block:
US-09-276-935b-14 x A1746915
Align seq 1/1 to: A1746915 from: 1 to: 794

1 LeuGIuValArgProLysGluSerTyrPasnHsAlaAspPheValHisGly 17
|||||.....
330 CTAGAGATGAGACCTGAGGAGAGCTGGAGCGAGTTGGCTTGATCACTG 379
17 sGluAspThrGlnSerValProGlyLysProSerValAsnAlaAspGluG 34
|||||.....
380 TGAACAGCAGACTCTGCTTGAAGAGCC...ATCAGCTAGAGGAGG 426
34 LuValGlyGlyProGlnIleCysArgValCysGlyAspLysAlaThrGly 50
|||||.....
427 AAGATGGAGGCTTTCGAATCTGCCGTATGTGGGACACAGCCCAATGGC 476
51 TyrHisPheAsnValMetThrCysGluGlyCysLysGlyPhePheArg 67
|||||.....
477 TACCACTTCATATGATATACCGTGTAAAGATGCAAGGGTTTTCACAG 526
67 gAlaMetLysArgAsnAlaArgLeuArgCysProPheArgLysGlyAlaC 84
|||||.....
527 GGCATGAACGCATGTCTCGCTAGCTGCCCTTCCGCAAGGAACT 576
84 yscGluIleThrArgLysThrArgArgGlnCysGlnAlaCysArgLeuArg 100
|||||.....
577 GCGAGATACCCCGAAGACACAGCAGCGAGTGCACGCCCTTGCGCG 626
101 LysCysLeuGlnSerCylMetLysGluMetIleMetSerAspGluAl 117
|||||.....
627 AAGTCCCTGAGAGTGGCATCAAGAAAGATGATATGTCATGCCCG 676
117 aValGluGluArgArgAlaLeuIleLysArgLysLysSerGluArgThrG 134
|||||.....
677 TGTGAGCAGCAGCGCGGCTTGATCAACAGAGAGAGGAGANNAGATTG 726
134 lYThrGlnProLeuGlyValGlnGlyLeuThrGlnGlnGlnArgMetMet 150
|||||.....
727 AGCTTCACCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 776
151 lLeuArgGlnLeuMetAsp 156
|||||.....
777 ATCCAGAGCTGATGAGC 794

seq_name: gb_est17:A1248626

seq_documentation_block: 343 bp mRNA EST 01-DEC-1998
LOCUS A1248626
DEFINITION qn77b12.x1 Soares_fetal_liver_spleen_INFUS_S1 Homo sapiens cDNA
clone IMAGE:1850687 3' similar to gb:J03258 VITAMIN D3 RECEPTOR
(HUMAN); mRNA sequence.
ACCESSION A1248626
VERSION A1248626.1 GI:3844023
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 343)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: rsaps-r@mail.nih.gov
This clone is available royalty-free through LNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 608 Std Error: 0.00
Seq primer: -400p from Gluco
High quality sequence stop: 338.
FEATURES
location/Qualifiers

```


610 AGSGCC 615

seq_name: qb_est43:AM158294

seq_documentation_block:

LOCUS AM158294 432 bp mRNA EST 05-NOV-1999

DEFINITION za39e07.x1 Xenopus EST library Xenopus laevis cDNA clone za39e07

5' mRNA sequence.

ACCESSION AM158294

VERSION AM158294.1 GI:6270323

KEYWORDS

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;

Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 432)

Schutz, K., de la Bastide, M., Huang, E.-N., Nascimento, L., Preston, R.,

Shah, R., Swaby, J., Shekter, M., Spiegel, L., Vil, M.D. and McCombie

JOURNAL

COMMENT

TITLE

Expressed sequence tags from Xenopus

Unpublished (1999)

Contact: W. Richard McCombie

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: za39 row: e column: 07

Seq primer: M13 universal forward primer

High quality sequence stop: 432.

FEATURES

source

1..432

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone_lib="Xenopus EST library"

/tissue_type="total brain tissue"

/cell_line="W2-TGA"

/dev_stage="tadpole"

/note="Vector: lambda zap 1; Site 1: Xba1; This library

was supplied by Holly Cline (Cold Spring Harbor labs).

CDNA synthesis with oligo df Xba 1 (Xba 1 cloning site).

RNA: stage 50-56 tadpoles, total brain tissue, GTC

extraction method."

BASE COUNT 112 a 90 c 142 g 88 t

ORIGIN

alignment_scores:

Quality: 389.50 length: 109

Ratio: 4.015 Gaps: 1

Percent Similarity: 88.991 Percent Identity: 64.220

alignment_block:

US-09-276-935B-14 x AM158294 ..

Align seq 1/1 to: AM158294 from: 1 to: 432

18 GUAAspThrGluSerValProGlyLysProSerValAsnAlaAspGlu 34

108 GAGGAGGACAGACGCCCTTAACAGTTGTGGGAGCGGGAAGACGAGGA 157

34 uValGlyGlyProGlnIleCysArqValCysGlyAspIysAlaThrGlyT 51

158 CGATGGGACCCCAAGATCTGCGTGGTGGGACCGGACCGACATGGGT 207

51 yTrHisPheAsnValMetThrCysGluGlyCysIysGlyPhePheArqArq 67

208 ATCACTCAATGCTATGACCTGCGAGGCGTCAAGGATTTCTACGCGG 257

68 AlMetLysArqAsnAlaArqLeuArqCysProPheArqLysGlyAlaC 84

|||||

256 GCGGTGAAGAGGAACTTGGCGCTCACTGCCCTTC...GAAATTGCTG 304

84 sCtUlleThrArqLysThrArqArqGlnCysGlnAlaCysArqLeuArq 101

305 CTTCACTCAACAGAGCAATCGGCCACACCTGCTGCGCTGACCA 354

101 yScCysLeuGlnSerGlyMetLysGlnMetIleMetSerAspGluAla 117

355 AATGCTGACATCCGATGAGAAAGAGTATGATGATGATGATGATGATG 404

118 ValGluGluArqArqAlaLeuIleLys 126

405 GTGGAACAGACAGACGACGCTAATTAG 431

seq_name: qb_est40:AV652474

seq_documentation_block:

LOCUS AV652474 637 bp mRNA EST 07-SEP-2000

DEFINITION AV652474 GLC Homo sapiens cDNA clone GICDAG03 3', mRNA sequence.

ACCESSION AV652474

VERSION AV652474.1 GI:9873488

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 637)

Qian, B., Wu, T., Huang, Q., Kang, R., Gao, X., Xu, Z., Xiao, H.,

Zeng, L., Xu, S., Gu, Y., Chen, Z., and Han, Z.

Homo sapiens cDNA clone

Unpublished (2000)

Contact: Zengqun Han

Chinese National Human Genome Center at Shanghai

351 Qiao Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801914 (ex. 45)

Fax: 86-21-50801922

Email: hanzq@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

1..637

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="GICDAG03"

/clone_lib="GLC"

/tissue_type="corresponding non cancerous liver tissue"

/dev_stage="Adult"

/lab_host="SOLR"

/note="Vector: pBluescript sk(-); Site 1: EcoRI; Site 2:

Xho1"

BASE COUNT 154 a 179 c 170 g 134 t

ORIGIN

alignment_scores:

Quality: 387.50 length: 274

Ratio: 2.533 Gaps: 6

Percent Similarity: 56.044 Percent Identity: 34.066

alignment_block:

US-09-276-935B-14 x AV652474 ..

Align seq 1/1 to: AV652474 from: 1 to: 637

18 GUAAspThrGluSerValProGlyLysProSer.....ValAs 30

30 GAGAACACAGACGCTGTGACAGCACCCACACGATGATGATGATG 79

30 nAlaAspGlnGluValGlyGlyProGlnIleCysArqValCysGlyAsp 47


```

80  TAAAGAAATGAACTT.....AGCAATCTGTGATGATATGCGGAGC 120
47  ysalatthrclytrhispheasvalmetthrcysgluclyscysgly 64
121 AACGCAATGCTACCTTAAATGCTGCTGATGATGATGATGATGATGAT 170
64  pth-phararararalmetlysarararalaaarararararararar 80
171 TTTCTTAAAGCAACAGTCACTAG..... 194
80  gylsrllyalacysgluclythararararararararararararar 97
195 .....AGGCGCCAGCTGACAGGCT 213
97  ysarararararararararararararararararararararar 113
214 GCAGATTCGAGAGAGCTTAGATGCTGCGATGAGCAAGACATGATCTG 263
114 SGTASpGluAlaValGluGluAlaGluAlaGluAlaGluAlaGluAla 140
264 TCGGCAAGACCTGCGATGCGATGCGAGCAAGCAAGCAAGCAAGCAAG 313
140  rgluararararararararararararararararararararar 147
414 AATGCAAAATCTGTGCA.....ATGATTAAGCAAGC 345
147  luarararararararararararararararararararararar 163
446 AAGCAATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 395
164  thrthrphosphoribosylasparagluoproglyvalleucoser 180
496 AATATCTTGAAAGATTGCTGAGTTAAACCTGAACTGATGCTGAT 445
180  rgllyrlysrleucoproglyserleucualaproserararararar 197
446  TATATCAATC.....CTGTGCTGACCTGAGCTCT 476
197  latlyrtrpsergluvalarararararararararararararar 213
476 ..... 476
214  gluhuarararararararararararararararararararar 240
476 ..... 476
240  psercylcylglysodalrphoserleucoproglysermetalaras 247
477 .....GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 510
247  cctthrcylmetpharararararararararararararararar 263
511 AATCTTAAAGCAACAGTCACTAG.....ATGATTAAGCAAGCCTG 560
264  pharararararararararararararararararararararar 280
561 TTTCTTAAAGCAACAGTCACTAG.....ATGATTAAGCAAGCCTG 610
280  apthararararararararararararararararararararar 286
611 TGTGCAATCTGTGCACTG..... 629

```

seq name: qb_cst 76:BE574696

seq document id: 1053 bp mRNA EST 15 AUG:2000

FEATURES

DEFINITION 60133441P2 NC1.CCAAT_Mam6 Mus musculus cDNA clone IMAGE:4710866 5'

VERSION BE574696.1 GI:9817416

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

1. 1053

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:3710866"

/clone_lib="NC1.CCAAT_Mam6"

/sex="female; virgin"

/tissue_type="inflicting ductal carcinoma"

/dev_stage="5 months"

/note="organ: mammary; Vector: pCMV SP6R10; Site 1: Salt; Site 2: Not; cloned unidirectional; primers: cloned; library constructed by Life Technologies; Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 444 a 362 c 274 g 173 t

ORIGIN

alignment _score:

Quality: 383.00 Length: 276

Ratio: 2.140 Gaps: 9

Percent similarity: 64.855 Percent identity: 45.507

alignment _block:

US 09 276 935b 14 x BE574696 ...

Align seq 1/1 for BE574696 from: 1 to: 1053

48 Proclithecysararararararararararararararararar 54

147 cctthrcylglysodalrphoserleucoproglysermetalaras 196

54 0valmetthrcysgluclyscysglyrphoserararararararar 71

197 cctthrcylglysodalrphoserleucoproglysermetalaras 246

71 rarasalar 87

247 cctthrcylglysodalrphoserleucoproglysermetalaras 293

88 Arar 104

294 AAGCAATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 343

104 0sercylmetlysararararararararararararararararar 120

344 0ATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 393

120 0ar 147

394 TAAAGCAATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 443

147 0ar 153

444 GTCTGAGCCGCAAG...CTGTCTGAGGACACAGACATCATCGCATC 490

154 LeuMetAspAlaGlnMetLysThrPheAspThrThrPheSerHisPheLeu 170
 491 CTCTGTCATGCCGCCACCAAGACCTACGACCCGACCTATGCCACTTCCG 540

170 SasuPheArgLeuPro...GlyValLeuSerSerGlyGlyGluLeuP 185
 541 GGACTTCGGCCCTCCATTGCTGCACAGCGTAAGTACAGGAGCTATCTTC 590

185 roLuserLeuGlnAlaProSerArgLugluAlaAlaLys... 198
 591 CAAGGCCCACTCACTGCTTCGAGAGACTCTCTCAACCTCATCT 640

199TrpSerGlnValArgLysAspMetLysSerLeuLeu 210
 641 TGTACACCCCTCAGCTGATGATGACCCGACCTTCCAGATGGC 690

210 SvalSerLeuGlnLeuArgLysLysSerValTyrPheSerValTyrLysP 227
 691 TGTGAATGACCAAGCTCCATGATGACCTTGTGACCTTGTGACCTTGTG 740

227 roProAlaAspSerGlyGlyGluLeuPheSerLeuLeuProHisMet 243
 741 CG.....CTATCCATGCATGTCACCT 763

244 AlaAspMetSer...ThrTyrMetPheLysGlyLeuLeuSerPheAlaLys 259
 764 GGCGGACCTCGGCGACCTACAGCTCCCAAGGACCCGCTTCCGAC 813

259 SvalLeuSerTyrPheArgPheArgPheLeuArgLysGlnLeuSerLeu 276
 814 GATATCCCTCGGCTCAAGGACCTACCTCTTGTAGAGACGACGAGGCGCG 863

276 eufysGlyAlaAlaPheGluLeuGlyGlnLeuArgPheAsnThrValPhe 292
 864 CCGAGGACCGCGCGGACGACGACGACGATGCCGCCAACACGACACTCAC 913

292 eAsnAlaGluThrGlyThrTrp 299
 914 CGGACGACCGCGCTCAACCTCG 935

seq_name: gb_est42:AW107536

seq_documentation_block: 601 bp mRNA EST 20-OCT-1999
 LOCUS AW107536 u191a06.y1 Sugano mouse kidney mRna Mus musculus cDNA clone
 DEFINITION IMAGE:2159410.57 similar to gb:J03258 VITAMIN D3 RECEPTOR (HUMAN);
 gb:015548 Mus musculus beta 2 thyroid hormone receptor (MOUSE);,
 mRNA sequence.

ACCESSION AW107536
 VERSION AW107536.1 GI:6078336
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 601)
 AUTHORs Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person
 B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
 E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterston,R. and Willson,R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Other ESTs: u191a06.x1
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1002582
 Seq primer: custom primer used
 High quality sequence stop: 502.
 Location/Qualifiers

FEATURES

source

1..601

/organism="Mus musculus"

/strain="C57BL"

/db_xref="taxon:10090"

/clone_image="2159410"

/sex="female"

/sex="male"

/sex="mouse"

/sex="kidney"

/sex="mRna"

/sex="adult"

/sex="Dhl08"

/sex="PME18s-FL3"

/sex="Site_1"

/sex="Site_2"

/sex="Site_3"

/sex="Site_4"

/sex="Site_5"

/sex="Site_6"

/sex="Site_7"

/sex="Site_8"

/sex="Site_9"

/sex="Site_10"

/sex="Site_11"

/sex="Site_12"

/sex="Site_13"

/sex="Site_14"

/sex="Site_15"

/sex="Site_16"

/sex="Site_17"

/sex="Site_18"

/sex="Site_19"

/sex="Site_20"

/sex="Site_21"

/sex="Site_22"

/sex="Site_23"

/sex="Site_24"

/sex="Site_25"

/sex="Site_26"

/sex="Site_27"

/sex="Site_28"

/sex="Site_29"

/sex="Site_30"

/sex="Site_31"

/sex="Site_32"

/sex="Site_33"

/sex="Site_34"

/sex="Site_35"

/sex="Site_36"

/sex="Site_37"

/sex="Site_38"

/sex="Site_39"

/sex="Site_40"

/sex="Site_41"

/sex="Site_42"

/sex="Site_43"

/sex="Site_44"

/sex="Site_45"

/sex="Site_46"

/sex="Site_47"

/sex="Site_48"

/sex="Site_49"

/sex="Site_50"

/sex="Site_51"

/sex="Site_52"

/sex="Site_53"

/sex="Site_54"

/sex="Site_55"

/sex="Site_56"

/sex="Site_57"

/sex="Site_58"

/sex="Site_59"

/sex="Site_60"

/sex="Site_61"

/sex="Site_62"

/sex="Site_63"

/sex="Site_64"

/sex="Site_65"

/sex="Site_66"

/sex="Site_67"

/sex="Site_68"

/sex="Site_69"

/sex="Site_70"

/sex="Site_71"

/sex="Site_72"

/sex="Site_73"

/sex="Site_74"

/sex="Site_75"

/sex="Site_76"

/sex="Site_77"

/sex="Site_78"

/sex="Site_79"

/sex="Site_80"

/sex="Site_81"

/sex="Site_82"

/sex="Site_83"

/sex="Site_84"

/sex="Site_85"

/sex="Site_86"

/sex="Site_87"

/sex="Site_88"

/sex="Site_89"

/sex="Site_90"

/sex="Site_91"

/sex="Site_92"

/sex="Site_93"

/sex="Site_94"

/sex="Site_95"

/sex="Site_96"

/sex="Site_97"

/sex="Site_98"

/sex="Site_99"

/sex="Site_100"

/sex="Site_101"

/sex="Site_102"

/sex="Site_103"

/sex="Site_104"

/sex="Site_105"

/sex="Site_106"

/sex="Site_107"

/sex="Site_108"

/sex="Site_109"

/sex="Site_110"

/sex="Site_111"

/sex="Site_112"

/sex="Site_113"

/sex="Site_114"

/sex="Site_115"

/sex="Site_116"

/sex="Site_117"

/sex="Site_118"

/sex="Site_119"

/sex="Site_120"

/sex="Site_121"

/sex="Site_122"

/sex="Site_123"

/sex="Site_124"

/sex="Site_125"

/sex="Site_126"

/sex="Site_127"

/sex="Site_128"

/sex="Site_129"

/sex="Site_130"

/sex="Site_131"

/sex="Site_132"

/sex="Site_133"

/sex="Site_134"

/sex="Site_135"

/sex="Site_136"

/sex="Site_137"

/sex="Site_138"

/sex="Site_139"

/sex="Site_140"

/sex="Site_141"

/sex="Site_142"

/sex="Site_143"

/sex="Site_144"

/sex="Site_145"

/sex="Site_146"

/sex="Site_147"

/sex="Site_148"

/sex="Site_149"

/sex="Site_150"

/sex="Site_151"

/sex="Site_152"

/sex="Site_153"

/sex="Site_154"

/sex="Site_155"

/sex="Site_156"

/sex="Site_157"

/sex="Site_158"

/sex="Site_159"

/sex="Site_160"

/sex="Site_161"

/sex="Site_162"

/sex="Site_163"

/sex="Site_164"

/sex="Site_165"

/sex="Site_166"

/sex="Site_167"

/sex="Site_168"

/sex="Site_169"

/sex="Site_170"

/sex="Site_171"

/sex="Site_172"

/sex="Site_173"

/sex="Site_174"

/sex="Site_175"

/sex="Site_176"

/sex="Site_177"

/sex="Site_178"

/sex="Site_179"

/sex="Site_180"

/sex="Site_181"

/sex="Site_182"

/sex="Site_183"

/sex="Site_184"

/sex="Site_185"

/sex="Site_186"

/sex="Site_187"

/sex="Site_188"

/sex="Site_189"

/sex="Site_190"

/sex="Site_191"

/sex="Site_192"

/sex="Site_193"

/sex="Site_194"

/sex="Site_195"

/sex="Site_196"

/sex="Site_197"

/sex="Site_198"

/sex="Site_199"

/sex="Site_200"

/sex="Site_201"

/sex="Site_202"

/sex="Site_203"

/sex="Site_204"

/sex="Site_205"

/sex="Site_206"

/sex="Site_207"

/sex="Site_208"

/sex="Site_209"

/sex="Site_210"

/sex="Site_211"

/sex="Site_212"

/sex="Site_213"

/sex="Site_214"

/sex="Site_215"

/sex="Site_216"

/sex="Site_217"

/sex="Site_218"

/sex="Site_219"

/sex="Site_220"

/sex="Site_221"

/sex="Site_222"

/sex="Site_223"

/sex="Site_224"

/sex="Site_225"

/sex="Site_226"

/sex="Site_227"

/sex="Site_228"

/sex="Site_229"

/sex="Site_230"

/sex="Site_231"

/sex="Site_232"

/sex="Site_233"

/sex="Site_234"

/sex="Site_235"

/sex="Site_236"

/sex="Site_237"

/sex="Site_238"

/sex="Site_239"

/sex="Site_240"

/sex="Site_241"

/sex="Site_242"

/sex="Site_243"

/sex="Site_244"

/sex="Site_245"

/sex="Site_246"

/sex="Site_247"

/sex="Site_248"

/sex="Site_249"

/sex="Site_250"

/sex="Site_251"

/sex="Site_252"

/sex="Site_253"

/sex="Site_254"

/sex="Site_255"

/sex="Site_256"

/sex="Site_257"

/sex="Site

9-70 The following are the first three terms of a geometric sequence:

Z1 Issue type: Tumor, biopsy sample

[illegible]


```

147 CTGAGATTTAGATCTCAAGCAAAATCTCAATCATCTGACTTTCTTCAACCTG 196
17 5815AAGTTTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 34
197 TACGATACACGATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 246
34 1041GAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTG 50
247 AAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 296
51 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 67
297 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 346
67 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 84
347 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 496
84 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 100
497 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 446
101 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 117
447 AAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 496
117 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 134
497 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 546
144 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 150
547 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 596
151 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 167
597 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 646
167 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 183
647 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 696
183 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 196
697 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 746
200 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 216
747 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 796
216 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 237
144 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 254
796 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 796

```

seq_name: /cgu2.6/ptcdat4/1/1m/5A.C00M.seq.us 07-737 736B 6

seq_documentation_block:

Sequence by: Application US/0773736B

Patent No. 5260199

GENERAL INFORMATION:

APPLICANT: Indiana, Hector F.

APPLICANT: Rossi, Troy K.

APPLICANT: Prabh, Jean M.

TITLE OF INVENTION: Method of Producing

TITLE OF INVENTION: 1,25-dihydroxyvitamin D3 Receptor Protein

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Carl K. Schwartz, Esq., c/o Quattrone & Brady

STREET: 411 East Wisconsin Avenue

CITY: Milwaukee

STATE: Wisconsin

COUNTRY: U.S.A.

ZIP: 53202

```

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/0773736B
FILING DATE: 19910740
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Schwartz, Carl K.
REGISTRATION NUMBER: 29,437
REFERENCE/BOOK NUMBER: 96-296-2185-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 414-277-5715
TELEFAX: 414-277-5774
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2043 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLGY: linear
MOLECULE TYPE: cDNA
HYPOHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rat
PUBLICATION INFORMATION:
AUTHORS: Burnstock, James K.
AUTHORS: Wiese, Russell J.
AUTHORS: Medina, No. 5260199yo
AUTHORS: Medina, Hector F.
TITLE: Structure and regulation of the rat
TITLE: 1,25-dihydroxyvitamin D3 receptor
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 85
PAGES: 9499-9502
DATE: December 1988
US-07-737-736B-6

```

alignment_scores:

Quality:	762.00	Length:	410
Ratio:	2.712	Gaps:	8
Percent Similarity:	68.547	Percent Identity:	40.000

alignment_block:

US-09-276-935b-14 x US 07-737-736B-6 ..

Align seq 1/1 to: US-07-737-736B-6 from: 1 to: 2043

```

38 1041GAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 54
155 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 204
54 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 71
205 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 254
71 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 87
255 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 401
87 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 104
402 AAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT 351
104 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 121
402 GATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 401
121 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 137

```



```

402 AGAGGAGATGATATGAGAGAAAGAGAGAGAGGCTTGAAGGACAGT 451
138 LeuGlyValGlnGlyLeuThrGluGluGlnArgMetCysLeuGlu 154
452 CTGAGGCCCAAG...CTATCTGAAGAAACACACACATCATACCATCT 498
154 uMetAspAlaGlnMetLysThrPheAspThrPheSerHisPheLys 171
499 GCTGGAGGCCCAACACAGACCTATGACCCCACTACCTGACTTAAAG 548
171 snPheArgLeuProGlyValLeu...SerSerGlyCysGluLeuPro 185
549 ACTTCGGAGCTCCAGTCTGATGAGAGGAGTACAGGAGCTATTCCTCA 598
186 GluSerLeuGlnAlaProSerArgGluGluAlaAlaLysTrpSerGln 202
599 AGG...CCACACCTGAGCTCTCCGGGAACTCCCTCCCTC 636
202 LatGlyAspLeuCysSerLeuLysVal... 211
637 CAGCTGAGCTGATACACACCTCAGTACATGATGAACCATCCGGCT 686
212 ...SerLeuGlnLeuArgGlyValAspGlySerValTrpAsnTyrLys 226
687 TTTCGAACCTGATCTGACGAGAGAGATTCGAT... 721
227 ProProAlaAspSerGlyGlyLysGluLeuPheSerLeuLeuProHis 243
722 GACCCGCTGATCTGACCTGACCTGCTCCTCTCTCCATGCTGCCCACT 771
243 tAlaAspMetSerThrTyrMetPheLysGlyLeuLeuSerPheAlaLys 260
772 GCGTGAAGCTGTCAGTACAGACATCCAAAGTTCATGCGCTTGGCAGA 821
260 allSerTyrPheArgAspLeuProIleGluAspGlnLeuSerLeuLeu 276
822 TGATCCAGATTCAGATTCACCTCCGATGACCGATTCGCTGCT 871
277 LysGlyAlaAlaPheGluLeuGlyGlnLeuArgPheAsnThrValPhe 293
872 AAGTCAAGGCCCATGAGTATCATGTATGCTCCACACCATCTTCAC 921
293 nAlaGluThrGlyThrProLysGlyArgLeuSerTyrCysLeuLys 310
922 CATGATGATATATCTCTGGAGCATCTGACGACGACGACGACGACG 971
310 spThr...AlaGlyGlyPheGlnGlnLeuLeuGluProMet 323
972 TCACCGATGCTCCAAAGCTGGGACACCTGAGCTGATCGACCCCTC 1021
324 LeuLysPheHisTyrMetLeuLysLysLeuGlnLeuHisGluGluTyr 340
1022 ATAAAGTTCCAGCTGGGCTGAAACACTGAACTTACATGAGGAAAGCA 1071
340 rValLeuMetGlnAlaIleSerLeuPheSerProAspArgProGlyVal 357
1072 TGCTCTTCTATGCGCATCTGCACTTCTCTCCCGACGACGACTGGGCT 1121
357 euGlnHisArgValValAspGlnLeuGlnGlnGlnPheAlaIleThrLeu 373
1122 AGGACGGCCAGAGCTGTGGAAAGCATTCAGAGCCGCTATCCACAGCCT 1171
374 LysSerTyrIleGlyCysAsnArgProGlnProAlaHisArgPheLeu 390
1172 CAGACCTACATCCGCTGCGGACCCGCGGACGACGACGACGACGCTCTA 1221
390 eleuLysIleMetAlaMetLeuThrGluPhe... 400
1222 TGCCAGATGATCCGAACCTGGCCGACCTGGGAGCTCAGCAGAGAAC 1271
400 ... 400
1272 ACTGCAACAAATACGCTCTCTCTCTTCAGGCGCGAGATAGCATGAG 1321

```

```

401 AlaThrProLeuMetGlnLeuLeuPheGly 410
1322 CTCACACCCCTTGTGTGTGAGAGTGTTCGC 1351
seq_name: /cgn2/6/ftcdala/1/fna/5A_COMB.seq;us-07-737-736b-5
seq_documentation_block:
? Sequence 5, Application US/07737736B
? Patent No. 5260199
? GENERAL INFORMATION:
? APPLICANT: Deluca, Hector F.
? APPLICANT: Ross, Troy K.
? APPLICANT: Prahl, Jean M.
? TITLE OF INVENTION: Method of Producing
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Carl R. Schwartz, Esq., c/o Quarles & Brady
? STREET: 411 East Wisconsin Avenue
? CITY: Milwaukee
? STATE: Wisconsin
? COUNTRY: U.S.A.
? ZIP: 53202
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/07/737,736B
? FILING DATE: 19910730
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Schwartz, Carl R.
? REGISTRATION NUMBER: 29,437
? REFERENCE/DOCKET NUMBER: 96-296-2185-2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 414-277-5715
? TELEFAX: 414-277-5774
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1399 base pairs
? TYPE: NUCLEIC ACID
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Homo sapiens
? PUBLICATION INFORMATION:
? AUTHORS: Baker, Andrew R.
? AUTHORS: McDonnell, Donald P.
? AUTHORS: Hughes, Mark
? AUTHORS: Crisp, Tracey M.
? AUTHORS: Mangelsdorf, David J.
? AUTHORS: Haussler, Mark R.
? AUTHORS: Pike, J. W.
? AUTHORS: Shine, John
? AUTHORS: O'Malley, Bert W.
? TITLE: Cloning and expression of full-length cDNA
? TITLE: encoding human vitamin D receptor
? JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
? VOLUME: 85
? PAGES: 3294-3298
? DATE: May-1988
? US-07-737-736B-5

```

```

alignment_scores:
? Quality: 75.00
? Ratio: 2.688
? Length: 412
? Gaps: 8

```


Percent Similarity: 67.718 Percent Identity: 40.291
 alignment block:
 US-09-276-935b-14 x US-07-747-746b-5 ...
 Align seq 1/1 for US-07-747-746b-5 from: 1 to: 1399

```

48 ProGlnIleCysAlaValCysGlyAspLysAlaThrGlyThrHisProAs 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 GCGGAGATCTGGGGTGTGTGTAAGAGCAAGCAACATGCTTCTACTTGA 225

54 MetThrCysGlnIleCysGlySerLysPheAsnValAlaMetLysA 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 TGGTATGATCTGGAGAGCTGCAAGAGCTCTCTCAAGAGCAAGATCAAG 275

71 GAGAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGL 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 GCAAGAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAG 422

88 ArgLysThrArgArgGlnCysGlnAlaCysArgAlaArgLysGlnAl 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 AAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 372

104 ArgGlnLysMetLysLysGlnMetLysMetSerAspGlnAlaValGln 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
373 GATGAGATATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422

121 GAGAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGLAAAGL 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
423 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 472

138 LeuGlyValLeuGlyLeuThrGlnGlnGlnAlaGlnMetLysArgAl 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
473 CTGCTGAGCTCAAG...CTGCTGAGCTCAAGCTCAAGCTCAAGCTCAAG 519

154 MetAspAlaGlnMetLysThrAspThrThrPheSerHisPheLysA 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 569

171 SerPheArgGlnProGlyValLeuSerSerGly...CysGln 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
570 AGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619

184 LeuArgGlnSerLeuGlnAlaProSerArgGlnAlaAlaLysTrpSe 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
620 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663

200 LeuValAlaValAspLysPheCys...S 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 713

208 GLeuLysValSerLeuGlnAlaArgGlyValAspGlySerValTrpAsn 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
714 GAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 754

225 TyrLysProAlaAspSerGlyCysGlnLeuPheSerLeuLeuPro 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 ...GAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 798

241 HisMetAlaAspMetSerThrTyrMetPheCysGlyLeuLeuSerPheA 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
799 GAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 848

258 LeuLysValLeuSerTyrPheAspLysProLeuGlnAspGlnHisSer 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
849 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898

275 LeuLeuLysValAlaAlaPheSerLeuArgCysGlnLeuArgPheAsn 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
899 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948

291 LysAsnAlaGlnThrGlyThrThrPheCysGlnArgGlnSerTrp... 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

949 CTTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
307 ...CysLeuGlnLysPheAlaGlyValPheGlnGlnLeuLeuLeuGln 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
999 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1048

422 ProMetLeuLysPheHisTyrMetLeuLysLysLeuLeuGlnHisGln 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1049 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1098

338 GlnLeuValLeuMetGlnAlaHisSerLeuPheSerProAspArgPro 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1099 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1148

455 LysValLeuLeuHisArgValValAspGlnLeuGlnGlnLeuPheAla 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1149 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1198

372 ThrLeuLysSerTyrLeuGlnCysAsnArgProGlnProAlaHisArg 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1199 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1248

388 LeuPheLeuLysLeuMetAlaMetLeuThrGlnPhe... 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1249 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1298

400 ... 400
1299 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1348

401 ...AlaThrProLeuMetGlnLeuLeuPheCys 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1349 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394

seq_name: /cruz_6/products/1/jna/5A_COMB.seq:US-08-459-489-1

seq_documentation_block:
? Sequence 1, Application US/08459489
? Patent No. 5686574
? GENERAL INFORMATION:
? APPLICANT: David D. Moore et al.
? TITLE OF INVENTION: CAR RECEPTORS AND RELATED
? NUMBER OF INVENTIONS: 10
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: Massachusetts
? COUNTRY: U.S.A.
? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5" Diskette, 1.44 MB
? OPERATING SYSTEM: IBM OS/2 Model 5.0z or 5.5x
? SOFTWARE: WordPerfect (Version 5.0)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/459,489
? FILING DATE:
? CLASSIFICATION: 455
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 07/841,450
? FILING DATE: February 26, 1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Paul T. Clark
? REGISTRATION NUMBER: 30,162
? REFERENCE/WORK NUMBER: 00786/126001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 542-5070
? TELEFAX: (617) 542-8906
? TELEX: 200154
? INFORMATION FOR SFO ID NO: 1:
? SPOUNGE CHARACTERISTICS:

```



```

172 PhcAqLpPheo6tlyValLeuSerSoclycysLeuPheo6tlyLeu 188
714 ..... 714
188 uGhAaLpPheo6tlyValLeuSerSoclycysLeuPheo6tlyLeu 205
715 ..... 715
205 sPheo6tlyValLeuSerSoclycysLeuPheo6tlyLeu 221
742 ..... 742
222 ValTPAsuTylystrGhAaLpPheo6tlyValLeuSerSoclycysLeu 248
754 ..... 754
248 LpPheo6tlyValLeuSerSoclycysLeuPheo6tlyValLeu 265
789 ..... 789
289 LpPheo6tlyValLeuSerSoclycysLeuPheo6tlyValLeu 301
940 AGcAaGhAaLpPheo6tlyValLeuSerSoclycysLeu 970
402 ..... 402
971 TGAAAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1020
417 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1070
1021 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1070
443 uGhAaLpPheo6tlyValLeuSerSoclycysLeuPheo6tlyLeu 450
1071 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1120
450 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1170
1121 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1170
467 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1170
1171 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1220
483 uGhAaLpPheo6tlyValLeuSerSoclycysLeuPheo6tlyLeu 498
1221 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1270
498 hT ..... 498
1271 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1320
402 ..... 402
1421 AGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1459

```

seq name: /v02-6/p10414/1/lna/93b_000b.seq:us_08_649_619b_1

seq documentation block:

```

1 Sequence 1, Application us/08649619b
2 Patient No. 5671916
3 GENERAL INFORMATION:
4 APPLICANT: NAKAMURA, YOSHIO
5 APPLICANT: SAITO, HIROKO
6 TITLE OF INVENTION: BURN PROTEIN AND DNA

```

```

1 TITLE OF INVENTION: ENCODING THE SAME
2 NUMBER OF SEQUENCES: 11
3 CORRESPONDENCE ADDRESS:
4 ADDRESS: FLYNN, THOMAS, JR.
5 ADDRESS: 2026 Kambling Road
6 CITY: Kalamazoo
7 STATE: Michigan
8 COUNTRY: USA
9 ZIP: 49008-1699
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb
12 MEDIUM TYPE: Storage
13 OPERATING SYSTEM: MS-DOS 5.0
14 SOFTWARE: Wordperfect 5.0
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/649,619b
17 FILING DATE:
18 CLASSIFICATION: 435
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: JP6-226270
21 FILING DATE: 21-SEPT-1994
22 APPLICATION NUMBER: PCT/JP95/01909
23 FILING DATE: 21-SEPT-1995
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Terry J. F. Chapman
26 REGISTRATION NUMBER: 42549
27 REFERENCE/EXCERPT NUMBER: Futura Case 144
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (616) 381-1156
30 TELEFAX: (616) 381-5465
31 INFORMATION FOR SEQ ID NO: 1:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 1979
34 TYPE: nucleic acid
35 STRANDEDNESS: double
36 TOPOLOGY: linear
37 MOLECULE TYPE: cDNA to mRNA
38 ORIGIN: SOURCE:
39 ORGANISM: Homo sapiens
40 IMMEDIATE SOURCE:
41 LIBRARY: Human mammary gland cDNA
42 FEATURE:
43 NAME/KEY: cOS
44 LOCATION: 206..1591
45 IDENTIFICATION METHOD: experimental examination
46 US-08-649-619b-1

```

```

1 alignment scores:
2 quality: 477.50 length: 497
3 percent similarity: 1.844 gaps: 18
4 percent identity: 52.113 percent identity: 26.761

```

alignment block:

```

1 US-09-276-935b-14 x US-08-649-619b-1
2 Align seq 1/1 to: US-08-649-619b-1 from: 1 to: 1979

```

```

1 5 pPheo6tlyValLeuSerSoclycysLeuPheo6tlyValLeuSerSoclycysLeu 19
2 11 ..... 11
3 256 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 405
4 19 pPheo6tlyValLeuSerSoclycysLeuPheo6tlyValLeuSerSoclycysLeu 43
5 406 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 355
6 44 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 46
7 456 GAGGAGATTCAGCTTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 405

```


1 LENGTH: 2040 base pairs
 2 TYPE: nucleotide
 3 STRANDNESS: single
 4 TOPOLOGY: linear
 5 MOLECULE TYPE: DNA (genomic)
 6 HYDROPHOBIC: No
 7 ANTI SENSE: No
 8 US 09 276 935b-14

alignment scores:

Query: 477..50 Length: 497
 Ratio: 1.844 Gaps: 18
 Percent Similarity: 52.114 Percent Identity: 26.761

alignment block:

US 09 276 935b-14 x US 09 276 935b-1

Align seq 1/1 to: US 09 276 935b-1 from: 1 to: 2040

5 ProctylusserTPAsnHisAlaAspHisValHisCysGlu.....AS 19
 111
 295 GTCCTGACAGCTGGAGCTGCTTCTTCTTCAACCACTGTAATACAGAGC 444
 19 pHisHisSerValProGly.....LysProSerValAsnAlaAspGlu 34
 111
 445 GTCCTGACAGCTGGAGCTGCTTCTTCTTCAACCACTGTAATACAGAGC 494
 44 GluValGly..... 46
 495 GAGGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 444
 47Glyp 48
 445 GTCCTGACAGCTGGAGCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 494
 48 TGGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 54
 495 ACGAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 544
 55 ValMetHisCysGluGlyCysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 71
 111
 545 GTCCTGACAGCTGGAGCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 594
 71 GAsnAlaValGluAlaGlyCysProHisArg.....LysGlyValAlaCysGlu 87
 111
 595 TGGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 644
 87 HisArgLysHisAlaGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 104
 111
 645 ACGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 694
 104 GluSerGlyMetLysLysGluMetLysMetSerAspGluAlaValGluGlu 120
 111
 695 GATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 744
 120 GATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 146
 111
 745 GAGGAGAT.....ATGCTGAGGAGAT..... 788
 147ProGluGlyValGluGly..... 142
 789 GATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 848
 142 142
 849 GATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 888
 143LeuThrGluGluGluGluGluGluGluGluGluGluGluGlu 155
 111
 889 GATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 948
 155 GAspAlaGluMetLysThrHisAspThrHisSerHisSerHisSerHis 171

976 976
 949 TGGCTGAGCAACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 976
 172 HisArgLysHisAlaGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 188
 976 976
 188 GCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 205
 977 1004
 205 SpHisCysSerLeuLysValSerLeuGluHisArgGlyGluAspGlySer 221
 111
 1004CTGAGGCTGAGAT..... 1015
 222 ValThrPheSerLysProHisAlaAspSerGlyGlyLysGluHisPheSer 248
 111
 1016CTGAGGCTGAGAT..... 1030
 248 HisLeuPheProHisMetAlaAspMetSerThrThrMetPheCysGlyLeu 255
 111
 1051 G.....CTGCTGACGAGCTGAGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1091
 255 HisThrAlaLysValHisSerThrPheArgAspLeuProLeuGluAsp 271
 111
 1092 TGGATCTGCTGAGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141
 272 GluHisLeuSerLeuLysGlyValHisAlaPheGluHisCysGluHisArgPhe 288
 111
 1142 GAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1191
 288 HisThrValHisAsnAlaGluThrGlyThrThrPheCys..... 301
 1192 ACGCTGAGCTGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCTCT 1242
 302GlyArgLeuSerThrCysLeuGluAsp.....ThrAlaGluGlyPheGlu 316
 1233 TCAAGAGCTTCACTACACCAAGAGGAGCTTCCAGCTGACAGGCTGAGAG 1282
 317 GluLeuLeuGluGluProMetLeuLysPheHisLysMetLeuLysLysLys 333
 1284 GTCATGATTCATCAACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1332
 333 HisLeuHisSerGluGluGluThrValLeuMetGluAlaHisLeuPheHis 350
 111
 1344 GAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1382
 350 GCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 366
 111
 1484 GAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1432
 367 GluHisPheAlaLeuThrLeuLysSerThrLeuLysAsnArgProGlu 383
 111
 1444 GAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1482
 383 HisPheAlaHisSerPhe.....LeuPheLeuLysHisMetAlaHisLeu 398
 111
 1484 GAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1542
 398 His.....GluPheAla..... 401
 1544 GAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1582
 402HisProLeuMetGluHisLeuPheGlyLeu 411
 1584 ACGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1621
 411

seq name: /var2/seqprod4/1/ins/5A 098b.seq US 09 276 935b-1
 seq document at: /var2/seqprod4/1/ins/5A 098b.seq US 09 276 935b-1
 Sequence 1: Application US 09 276 935b-1
 Patient No: 5679518
 GENERAL INFORMATION:


```

APPLICANT: Friedman, Ritan
APPLICANT: Holloway, M. Katharine
APPLICANT: Kodan, Gideon
APPLICANT: Rutledge, Su Jane
APPLICANT: Schmidt, Azriel
APPLICANT: Vogel, Robert
TITLE OF INVENTION: METHOD FOR FINDING RECEPTOR POTENTIATORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Merck & Co., Inc.
STREET: 126 East Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,283
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Dolan, Catherine A.
REGISTRATION NUMBER: 36,502
REFERENCE/DOCKET NUMBER: 19327
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4283
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ. ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2030 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-330-283-1

alignment_scores:
quality: 477.50 length: 497
ratio: 1.844 gaps: 18
percent similarity: 52.113 percent identity: 26.761

alignment_block:
US-09-276-935b-14 x US-08-330-283-1 ..
Align seq 1/1 to: US-08-330-283-1 from: 1 to: 2030

5 ProlysgluserTPRasnHisAlaAspPheValHisCysglu.....As 19
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 CCCCCCTCAGCTGGGCCCCCTTCTTTCACCCACTGTATAGGAGGAGG 344
19 pThlGluserval.Progly.....LysProserValAsnAlaAspqlu 33
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
345 GTCCGAGACCTGGCCGGGGGTCCGACCCCTGATGTCCAGCGACACTGAT 394
34 GluValGly..... 36
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
395 GAGGCCAGCTAGCCTCAGACAGCACTGGCTCATCCCAATCCCGAGA 444
37 .....GlyP 38
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
445 GGAACCGAGGCGCAAGAAAGAGGCCAGGCCCAACATGCTGGGCC 494
38 roGlnlleCysArgValCysGlyAspLysAlaThrGlyTyrHisPheasn 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
495 ACGAGCTTGCGCTGCTGTGGGACACAAGGCTCCGGCTTCACACTAGAAC 544

```

```

55 ValMetThrCysGluGlyCysLysGlyPhePheArgArgAlaMetLysAr 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
545 GTGCTCACTCCCAAGGCTGCAGAGGCTTCTCCGGCGAGTGTGCTCG 594
71 qasAlaArGluCysProPheArg...LysValCysGluIleT 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
595 TGTGTGGGCGAGCGGCTATGCCCTGGGGGTGGCGCAACTGCGAATGG 644
87 hrArgLysThrArgArgGlnGlnAlaCysArgLeuArgLysLysLeu 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
645 AGCCTTCATGCCGCAAGTGCACAGCTGCCGCTCGCAACTGCAG 694
104 GlusergLysMetLysLysGluMetIleMetSerAspGlnAlaValIleu 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
695 GAGCGAGGATGAGGAGGAGCTGCTCTTCTGGAAGAACATCTCGGAA 744
120 uArgArgAlaLeuIleLysArgLysLysSerGluArgThrGlyThrGln 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
745 GAAGAG...ATTCGGAACAGCAGCAGCAGAGTCAAGTACAGCT 788
137 .....ProLeuGlyValGlnGly..... 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
789 CGCAGTCACTGTGGGCGCGCAGGCGCAGCAGCTCAAGCTCTGGCCT 838
142 ..... 142
839 CGGGCTTCCCTGCTGAGCTGAGCGAGCGAGCGAGGCTCCGGGAGG 888
143 .....LeuThrGluGluGluArgMetLysIleArgGluLeu 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
889 CGAGGCTGTCCACTACAGCGCGCTCAGACACTAATGATCGACACTTGG 948
155 eAspAlaGluMetLysThrPheAspThrThrPheSerHisPheLysasn 171
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
939 TGGCGGCCCACTGACGTCAGCAACAAGCGCTCTTCC..... 976
172 PheArgLeuProGlyValLeuSerSerGlyGlyGluProGluSerLe 188
976 ..... 976
188 uGlnAlaProSerArgGluGlnAlaAlaLysTrpSerGlnValArgLysA 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
977 .....GACCAAGCCCAAGTCAAGCGCTGGCC..... 1003
205 sPheCysSerLeuLysValSerLeuGlnLeuArgLysAspLysSer 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1004 .....CTGGCGCCATAC..... 1015
222 ValTPRasnTyrTrpProAlaAspSerGlyGlyLysGluIlePheSe 218
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1016 .....CCCCAGTCCGAGATGCCCGCCCAACAGCTTTC 1050
238 rLeuLeuProHisMetAlaAspMetSerThrTyrMetPheLysGlyIle 255
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1051 C.....CACTTCAGSAGCTGGCATCATCTCAGTCCAGAGATATG 1091
255 LeSerPheAlaLysValIleSerTyrPheArgAspLeuProIleGluAsp 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1092 TGCACTGCTTAAAGCAAGTGTGTTCTCTGCGAGCGGGCGAGAGAG 1141
272 GlnIleSerLeuLeuLysGlyAlaAlaPheGluLeuGlyGlnLeuArgP 288
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1142 CAGATGCCCTCCTGAAGGCATCCACTATCGAGATATGCTGTAGAGAG 1191
288 eAsnThrValPheAsnAlaGluThrGlyThrTrpGluCys..... 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1192 AGCCAGGCGCTACACACAGACACA.....GAGTGTATACCTCT 1232
302 .....GlyArgLeuSerTyrCysLeuGluAsp...ThrAlaCysLysPheGln 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1233 TGAAGGACTTCAGCTACAGCAGAGGAGCACTTCCAGCTGAGGCTGCA 1282
317 GlnLeuLeuLeuGluIleProMetLeuLysPheHisTyrMetLeuLysLys 343

```



```

1  APPLICANT: Vogel, Robert
2  TITLE OF INVENTION: METHOD FOR FINDING RECEPTOR POTENTIAL
3  NUMBER OF SEQUENCES: 5
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Morck & Co., Inc.
6  STREET: 126 East Lincoln Avenue
7  CITY: Rahway
8  STATE: New Jersey
9  COUNTRY: US
10 ZIP: 07065-0907
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: Patent Release #1.0, Version #1.25
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: PCT/US95/13924
18 FILING DATE:
19 CLASSIFICATION:
20 ATTORNEY/AGENT INFORMATION:
21 NAME: Quagliato, Carol S.
22 REGISTRATION NUMBER: 35,330
23 REFERENCE/BOOKET NUMBER: 19327 PCI
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: (908) 594-3809
26 TELEFAX: (908) 594-4720
27 INFORMATION FOR SEQ. ID NO: 1:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 2030 base pairs
30 TYPE: nucleic acid
31 STRANDEDNESS: single
32 TOPOLOGY: linear
33 MOLECULE TYPE: DNA (genomic)
34 HYPOTHETICAL: NO
35 ANTI-SENSE: NO
36 PCT-US95-13924-1
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
97
```


alignment_block:
US-09-276-935B-14 x PCT-US95-13931-1

Align seg 1/1 to: PCT-US95-13931-1 from: 1 to: 2030

5	ProlysgluSerThrPasnH1	salalaspheValHisCysGlu.....As	19
295	CCCCCTCAGCCTGCGCGCCCTTTCTTCTTCACGCCACTGMAAGCGAGCG		344
19	PhrIuInsErval.ProcIy.....LysProSerValAsnAlaAspIu		33
345	GTCGGACCCCGGTGGCGCGGGGGGCCGCCAGCCGTGATGTCGCCAGCTCAT		394
34	GIuValGly.....		36
395	GAGGCGAGCTCAGCTGCGAGCAGACAGCTGGGTGCATCCCAATCCCGAGAA		444
37GlyP		38
445	GGAAACCAGAGCCCAAGCAAAAGGAGGCGCCAGCCCGCAAGATGCTGGGCC		494
38	ProGlnIleCysArrValCysGlyAspLysAlaThrGlyThrHisPheAsn		54
495	ACGACCTTGCCCTGTCTGTGTGGAGCAAGGCGCTCCGGGTTCACATACAG		544
55	ValMetIrrCysGluGlyCysLysGlyPhePheArqValAlaMetLysAr		71
545	GTCGTCAGCTCGGAAGGCTGCAGAGGCTTCTTCCGGCCACATGTGGTCG		594
71	qAsnAlaIrrAlaIrrArqCysProPheArq.....LysGlyAlaCysIuIleT		87
595	TGCTGGGGCCAGCGCTATGCTGCGGGGGTGGGGAACCTGCCAGATCG		644
87	hArIrrLysThrArqArqGlnCysGlnAlaCysArqLeuArqLysCysLeu		103
645	ACGCTTTCATGGGGGCAAGTCCAGCATGTCGGGCTCGCAAGTGCAG		694
104	GIuserGlyMetLysGlyGlnMetIleMetSerAspGluAlaValGluG		120
695	GAGCGAGGATGAGGAGGAGTGGCTCTTCTTGAGAAACAGATCGGAA		744
120	uArqArqAlaLeuIleLysArqLysLysSerGluArqThrGlyThrGln		136
745	GAAAGAG.....ATTCCGAAACACACACACACAGATCAGAGTCCACAGT		788
137ProLeuGIalGlnGly.....		142
789	GCGAGTCACTGTGGGCGCGCGAGGCGTACACAGTCAAGCTCTGGGCT		838
142		142
839	GGGGCTTCCCTGGTGGATCTGATGCGAGGCAAGCGAGGCTCCGGGGAAG		888
143LeuThrGluGlnGlarGmetMetIleArqIuLeuM		155
889	CGAGGCTGCCAGCTTACACAGCGGCTCCAGAAATATATGATCCAGCAGTTGG		938
155	eTAspAlaGIuMetLysThrPheAspPThrThrPheSerHisPheLysAsn		171
939	TGGCGGCCCAACGTGACATGCACAAACGCTCTCTTC		976
172	PheArqLeuProGlyValLeuSerSerIrrCysGluLeuProGlnSerLe		188
976		976
188	uGlnAlaProSerArqGluGluAlaAlaLysTrpSerGlnValaIrrLysA		205
977GACCAAGCCCAAGATCAGCGCTGGCGCC		1005
205	sPluEncysSerLeuLysValSerIleuGlnLeuArqGlyAlaSpIrrSer		221
1004CTGGGCGCCAGAC.....		1015

[illegible]

